10018929 .122101

FORM PTO-1390 (REV. 9-2001) U S DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

ATTORNEY DOCKET: S-31005A

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

U S APPLICATION NO. (If known, see 37 CFR 1.5: 10/018929

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/EP00/05761	June 21, 2000	June 23, 1999

TITLE OF INVENTION
Gene Involved in Epigenetic Gene Silencing

APPLICANT(S) FOR DO/EO/US
Yoshiki Habu et al.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. [ ] This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. [X] This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.
4. [ ] The US has been elected by the expiration of 19 months from the priority date (Article 31).
5. [X] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   - a. [X] is attached hereto (required only if not communicated by the International Bureau).
   - b. [ ] has been communicated by the International Bureau.
   - c. [ ] is not required, as the application was filed in the United States Receiving Office (RO/US).
6. [ ] An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
   - a. [ ] is attached hereto.
   - b. [ ] has been previously submitted under 35 U.S.C. 154(d)(4).
7. [X] Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
   - a. [X] are attached hereto (required only if not communicated by the International Bureau).
   - b. [ ] have been communicated by the International Bureau.
   - c. [ ] have not been made; however, the time limit for making such amendments has NOT expired.
   - d. [ ] have not been made and will not be made.
8. [ ] An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. [X] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. [ ] An English lanugage translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 20 below concern document(s) or information included:

11. [ ] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. [X] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. [X] A FIRST preliminary amendment.
14. [ ] A SECOND or SUBSEQUENT preliminary amendment.
15. [ ] A substitute specification.
16. [ ] A change of power of attorney and/or address letter.
17. [X] A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.
18. [ ] A second copy of the published international application under 35 U.S.C. 154(d)(4).
19. [ ] A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).
20. [X] Other items or information: Certified Copy of Priority Document GB 9914623.5

U.S. APPLICATION NO. (if known, see 37 CFR 1.5) 10/018929 | INTERNATIONAL APPLICATION NO

10018929 .122101

21. [x] The following fees are submitted:

BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):

	CALCULATIONS	PTO USE ONLY
Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO **$1040.00**		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO **$890.00**		
International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO **$740.00**		
International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) **$710.00**		
International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) **$100.00**		
ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 890.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	$	
Total claims	16 - 20 =		x **$18.00**	$	
Independent claims	2 - 3 =		x **$84.00**	$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ **$280.00**	$	

TOTAL OF ABOVE CALCULATIONS =	$	
[ ] Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2. +	$	
SUBTOTAL =	$	
Processing fee of **$130.00** for furnishing the English translation later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(f)).	$	
TOTAL NATIONAL FEE =	$	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +	$40.00	
TOTAL FEES ENCLOSED =	$ 930.00	
	Amount to be refunded:	$
	charged:	$

a. [ ] A check in the amount of $ ____________ to cover the above fees is enclosed.

b. [ ] Please charge my Deposit Account No. ____________ in the amount of $ __________ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. [X] The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 50-1744. A duplicate copy of this sheet is enclosed.

d. [x] Fees are to be charged to a credit card. **WARNING:** Information on this form may become public. **Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Customer Number 22847

Marcia R. Morton
SIGNATURE

Marcia R. Morton
NAME

46,942
REGISTRATION NUMBER

10018929 . 122101

531 Rec'd PCT/PTO 21 DEC 2001

FILING BY "EXPRESS MAIL" UNDER 37 § C.F.R. 1.10

I hereby certify that the following correspondence is being deposited with the United States Postal Service as "Express Mail Post Office to Addressee" in an envelope addressed to: BOX PCT, U.S. Patent and Trademark Office, P.O. Box 2327, Arlington VA, 22202 under Express Mail Label No. ET327548859US on December 21, 2001.

1) Transmittal Letter to the United States Designated/Elected Office (DO/EO/US) Concerning a Filing Under 35 U.S.C. 371 (FORM PTO-1390)

2) Copy of International Application As Filed

3) Amendment to the Claims of the International Application under PCT Article 19

4) Certified Copy of Priority Document GB 9914623.5

5) Recordation Form Cover Sheet

6) Assignment (2 sheets total)

7) Declaration and Power of Attorney for United States Application (2 total)

8) First Preliminary Amendment

9) A Computer-Readable Form and Paper Copy of the Sequence Listing and Statement of Verification

10) Credit Card Payment Form

Melissa Hardy
Name

Melissa Hardy
Signature

10018929 .122101

10/018929

531 Rec'd PCT/P 21 DEC 2001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	
Habu *et al.*	Art Unit: TBA
Serial Number: TBA	Examiner: TBA
Filed: December 21, 2001	Atty Docket: S-31005A
For: Gene Involved in Epigenetic Gene Silencing	

PRELIMINARY AMENDMENT

BOX PCT
U.S. Patent and Trademark Office
P.O. Box 2327
Arlington, VA 2202

Sir:

Applicants respectfully request that the above-captioned application be amended as follows in advance of prosecution:

IN THE SPECIFICATION:

At page 1, after the title, please insert:

This application is a § 371 of International Application No. PCT /EP00/05761, filed June 21, 2000.

IN THE CLAIMS

Please amend claim 8 as follows:

8. (Amended) The protein encoded by the open reading frame of claim 1.

Please add the following new claims:

11. (New) The protein encoded by the open reading frame of claim 2.

12. (New) The protein encoded by the open reading frame of claim 3.

13. (New) The protein encoded by the open reading frame of claim 4.

14. (New) The protein encoded by the open reading frame of claim 5.

15. (New) The protein encoded by the open reading frame of claim 6.

16. (New) The protein encoded by the open reading frame of claim 7.

REMARKS

The claims have been amended to eliminate multiple dependency and to ensure that they are in compliance with the rules of U.S. practice. No new matter has been added by these amendments.

If in the opinion of the Examiner, a telephone conference would expedite the prosecution of the subject application, the Examiner is invited to call the undersigned attorney.

The Commissioner has been authorized to charge the fees associated with the filing of the application which this amendment accompanies and associated papers to Applicant's credit card. If there are any additional fees required, authorization is hereby provided to charge such fees to Applicant's Deposit Account No. 50-1744 (in the name of Syngenta Biotechnology, Inc.).

Respectfully submitted,

Marcia R. Morton

Marcia R. Morton
Attorney for Applicants
Reg. No. 46,942

Syngenta Biotechnology, Inc.
Patent Department
P.O. Box 12257
Research Triangle Park, NC 27709-2257
(919) 541-8566
December 21, 2001

Version With Markings To Show Changes Made

The claims have been amended as follows:

8. (Amended) The protein encoded by the open reading frame of claim 1 [any one of claims 1 to 7].

--11. (New) The protein encoded by the open reading frame of claim 2.

12. (New) The protein encoded by the open reading frame of claim 3.

13. (New) The protein encoded by the open reading frame of claim 4.

14. (New) The protein encoded by the open reading frame of claim 5.

15. (New) The protein encoded by the open reading frame of claim 6.

16. (New) The protein encoded by the open reading frame of claim 7.

10018929 .122101

10/018929 S-31005A

531 Rec'd PCT 21 DEC 2001

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF

Habu *et al.*

SERIAL NO: TBA

FILED: Even Date Herewith

FOR: Gene Involved in Epigenetic Gene Silencing

ART UNIT: TBA

EXAMINER: TBA

BOX PCT
U.S. Patent and Trademark Office
P.O. Box 2327
Arlington, VA 22202

SUBMISSION OF SEQUENCE LISTING
INCLUDING STATEMENT OF VERIFICATION

Sir:

Applicant hereby provides a Computer Readable Form of the Sequence Listing as well as the Paper Copy thereof. The undersigned states that the Paper Copy and the Computer Readable Form, submitted in accordance with 37 CFR §1.821(c) and (e), respectively, are the same.

Respectfully submitted,

Marcia R. Morton

Marcia R. Morton
Attorney for Applicants
Reg. No. 46,942
(919) 541-8566

Syngenta Biotechnology, Inc.
Patent Department
P.O. Box 12257
Research Triangle Park, NC 27709-2257

Date: December 21, 2001

SEQUENCE LISTING

<110> Novartis AG
Novartis Research Foundation

<120> Gene involved in epigenetic gene silencing

<130> S-31005A

<140>
<141>

<150> GB 9914623.5
<151> 1999-06-23

<160> 33

<170> PatentIn Ver. 2.1

<210> 1
<211> 10329
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> intron
<222> (1009)..(1295)

<220>
<221> intron
<222> (2551)..(2673)

<220>
<221> intron
<222> (2753)..(2867)

<220>
<221> intron
<222> (3114)..(3506)

<220>
<221> intron
<222> (3681)..(3973)

<220>
<221> intron
<222> (4896)..(4975)

<220>
<221> intron
<222> (5218)..(5777)

<220>
<221> intron

<222> (5883)..(6082)

<220>
<221> intron
<222> (7481)..(7615)

<220>
<221> intron
<222> (7772)..(7914)

<220>
<221> intron
<222> (8071)..(8153)

<220>
<221> intron
<222> (8319)..(8451)

<220>
<221> intron
<222> (8630)..(8718)

<220>
<221> intron
<222> (8919)..(9000)

<220>
<221> intron
<222> (9212)..(9284)

<400> 1
aatatttaag tttggtttat attctttcta gtaatctttg aaatattgta agagataatg 60
cttctaataa ataacattgg atttattgga attaatgtat tgaaaaaact atgcaaatac 120
tacagtgtat tttggaacga ccaaaatgat atatgtaaac tttcgttcta gtcttctaca 180
tagtgtaata ggatagcgga caaggttgat cgactctaaa cattatgggt acgtaattcc 240
gcagtggtta cagtctactg tcgaggccaa actggtaatt aaacgtttga agtttagaga 300
aatattttga tgatgagtac cacaatcaaa gatgataggt gttaatcact gtaaaaatgt 360
tgattgaata ctacgaatgc agaacatata catattttta atctctttgg aatttttgtt 420
tttgttttta tcatttttga atacacgaag agctcagtta tatttcatat tgtatatgaa 480
tttgttctat ttaatcttca attctagcaa catactctta tgctaattcg tttcatattt 540
tagtatagta taaaaattac aaatttcaaa acaaactata agtaatatac taacatagtc 600
ggtgtaacat ttcgttaatt tcacataaca tatgttaatt acatatgtac actatttttg 660
aagtatttta taacttaaaa tatataaatt taaatctaag aaatcacaag catgagtttt 720
tccttccggt aatcgtaaaa tcaaaaatcg ctcgctcgag aaacgccggt gctagaagag 780
gaaagtaccg tacataatcc tgcgaaccca attctcgtct tcttcaaact cagttttccg 840
aaaccccaaa caccgcgagg attgcatggc ctgaagaacc acttaatcga gaattgtgct 900
ggaattctca aattttccct cgcgtttttc tttcacactc tcggaatcgg aaatttccac 960
caagctccgt caagcgatag attctgacaa ttacacactt tcgcgcaggt atgcttcctt 1020
ccctgtttta ggttggtgtt aatctatcgg tgaatcgaag gttttgggcc tcgggctttg 1080
cgttttaggt ttttcagaga atcttatcta cttggggatg gatcttaggc gtttgttaga 1140
tgtaactcat tagttttgca tataggaatt ttgatttgaa agttaggtcg ccggatttgt 1200
agacattttg tttgatggtc ttcttcggtg ctcacattct ttgtttttaa gtgcttgatt 1260
tggttgctaa ggtcctttcc gttgcgtgct ctcagtgaat atgaagaaag atgaaaagat 1320
tggtttgacg gggagaacca tttacaccag atccctagca gcttcaattc ctgcctcagt 1380
tgaacaagaa acccctggtt tgaggaggtc aagccggggg acaccatcta cgaaggtaat 1440

aactccagct tctgctacta gaaagtcaga gagactggct ccctcacctg cttcagtttc 1500
aaaaaagtcc ggtggaatcg tcaagaattc cacaccaagt tctttgcgaa ggtccaatag 1560
ggggaagact gaagtatcct tgcagagttc caaaggatca gataattcta tcaggaaagg 1620
agatacttca ccggatattg agcagagaaa ggatagtgtt gaagagtcga cagataagat 1680
caagcctata atgtcagccc gaagttacag ggcattgttt agagggaagc tcaaggaatc 1740
tgaggcatta gttgatgctt ccccaaatga agaggaacta gtagttgttg gttgttctcg 1800
ccgcatacct gcaggcaatg atgatgttca aggtaaaaca gattgtccac cacctgcaga 1860
tgcaggatca aaaaggctgc cagttgacga aactagtttg gacaagggca ctgattttcc 1920
tttgaaatca gttacggaga ccgagaagat agtgcttgat gcatccccca tagttgaaac 1980
tggggatgac agtgttatag gttcaccatc tgagaattta gagacacaaa agcttcaaga 2040
tggtaagaca gattgttcac cacctgcaaa tgcagaatcg aaaacgctgc cagttggtga 2100
aactagttta gaaaaagaat atccacaaaa gtttcaagat gataacacag attgtctacc 2160
acctgcaaat gcagaatcaa aaaggctgcc agttggcgaa actagtttag aaaaggacac 2220
tgattttcct ttgaaatcaa ctacggagac tggaaagatg gttctttatg catcccccat 2280
agttgaaact agggatgaca gcgttatatg ttcaccatct acaaatttag aaacccaaaa 2340
gcttcttgtc agtaaaactg gcttagaaac cgacatagtt ttgcctttga aaagaaaaag 2400
agacactgca gaaattgagc tggatgcatg tgctacagtt gcaaatggag atgatcacgt 2460
tatgagttct gatggggtca ttccatctcc atctgggtgc aaaaatgata atcgacctga 2520
aatgtgcaac acgtgtaaaa aacggcaaaa gtaagagttt ttttagtgtt gtctgtctat 2580
tgaaacgatc tgccaatgtt gaatgttggg cagatgggtt tgattcttag gatatatgtt 2640
ctgtattgta atgagttgtt caaaattttg aagggtcaac ggtgattgtc aaaataggag 2700
tgtttgctcc tgcattgtcc agccagttga agaatctgat aacgtgacac aggttggttt 2760
ctaattactt tcggagaccc gttaatcagt ggactcttaa atagttagat actagattta 2820
cttatccttt tacttgtaat ctgcaattct attttgcatt tgattaggat atgaaagaaa 2880
ctggaccagt tacgagcaga gaatatgagg agaacgggca aatacaacat ggtaaatcaa 2940
gtgatcccaa attctattct tcggtgtacc cagagtattg ggttcctgtg cagctatcag 3000
atgtacagct ggagcaatac tgtcagactc tcttctccaa atccttatct ctttcttcac 3060
tttcgaagat tgatcttgga gctctagaag aaactctcaa ttctgtaaga aaagtaagtt 3120
acttgatttt aaaaacactt attcttcaat gcacttgtga gttaagtacc cagttattac 3180
tggtgataag ataaagaaag caatagaaaa attgataagg tgttcaccgc attgcagcca 3240
aaaaaacaat tctgtgttcc atgctttcaa gaggttgtca cataggtgtt atgcctttct 3300
gtttgatgtt tggtagagca aaggttttgg gtctatttgt tttatgcttt tttgaaacac 3360
atagaacctg gcaaacttga cagttttggg gttgcttaga tatacgacta ttgtcggtca 3420
gcatcacatt ttctcaaggc ctctttctgc atgttaatgt gtgaatatat taaaatcttc 3480
tttatgtgtt tgcaacttgt tgacagacct gtgaccatcc atacgttatg gatgcatctt 3540
tgaaacaact gctcaccaag aatctggagt tgcatgaaat cctggatgta gaaattaaag 3600
cgagcgggaa acttcacctc cttgataaaa tgcttactca tataaaaaag aatggtttaa 3660
aagcagttgt cttctaccag gtgcattttc tattacttgc gaatgtgaat agctctatgt 3720
ttgtcatgaa tacgtcactt tgtgcattct caatatatgt gcattttctt tttgacaatg 3780
gaattctgtc ttgtattgaa atttgagtgg gatgaaagta tgctttttat cgtgcaatta 3840
tgaagtgtaa gttagccttc agcagtcagc tagcattatg agatatgctg aactaaaatg 3900
tttcttttct cttctttctt tttcgttata tgtgcctcat gtatgtttga attacagttt 3960
ttattttcag caggcaacac aaacccctga agggcttctg cttggtaata ttctcgaaga 4020
ttttgtgggt caaagatttg gtccaaaatc ttatgagcat gggatatatt cctcaaagaa 4080
gaactccgct ataaacaatt tcaacaagga gagtcaatgc tgtgttctgc tgttggaaac 4140
acgtgcctgc agtcaaacca ttaaactctt gcgagctgat gcgtttattc tttttggaag 4200
cagcttgaat ccatcgcatg atgttaagca cgtagagaag ataaaaatcg agtcatgttc 4260
tgaaagaact aagatattcc gattgtactc agtatgtaca gttgaagaaa aagccctgat 4320
tctggctagg caaaatatgc ggcaaaataa agctgtagag aacctaaacc gctctctcac 4380
gcacgcactg ctcatgtggg gggcgtcata cttatttgat aaactggatc attttcacag 4440
cagtgaaact ccagattcag gagtttcatt tgaacaatct attatggacg gcgtgattca 4500
tgaattctcg tccatacttt cttccaaagg tggagaagaa aatgaagtca agctgtgtct 4560
acttttggag gccaagcatg ctcagggaac ttacagcagt gattctactc tatttggtga 4620
agaccatatt aagttgtcag atgaagagag tccaaatata ttttggtcaa agctgttggg 4680
gggaaaaaat cctatgtgga aatacccttc agatactccc caaaggaatc gaaaacgagt 4740

```
tcagtatttt gagggttctg aagcgagtcc caaaactggc gatggtggaa atgcaaagaa 4800
gcgaaagaag gcttctgatg atgtcactga tccccgggtc actgatccgc cagtagatga 4860
tgatgaaaga aaggcctctg ggaaggatca catgggtaaa atagtttaat ttctgctccg 4920
atacctctag tgttcattga ttatgcaact actttgctga ctatctttcc tacaggggct 4980
ttggagtcac caaaagtcat aacactccag tcatcatgta aatcttctgg tacagatggt 5040
acattggatg gaaatgatgc ttttggcttg tattctatgg gcagccatat ctctggaatc 5100
ccagaggata tgttagctag tcaagattgg gggaaaatac cggatgaatc acagaggagg 5160
ctccacactg ttttaaagcc gaagatggca aaactttgcc aagttttgca tctttcagta 5220
agtggccttt ttcacctcca caacttattt tagccttgca tatgcttata tatagctgat 5280
tgcaactgta gttgttacct gatttcctgt tacagccaaa tgtgagagtt ttattcttca 5340
actatatcca tccgtttaag catattttat ttcttatatc tggcttcgtt accaatgcac 5400
tgttaaaatg agcaactgct gcacaaaaca gtaggtagtt atgtgcctca tgtcattcat 5460
tgtttattga agcaaagaaa tttctgtcta ctttacatga tccatctgtg ggagtatata 5520
actatatata accttaggcc tttgtacctg gctgatcaaa gacatgtcaa aagtttatct 5580
gttcgctgtt ggtatagaaa ctaatacagt gtctgatgct attttaaggt agtcttatgt 5640
cttcacatat tggctaatag atgtttccgc tgtcgtgtcc atatacttct gtgattatca 5700
cggtgctccg tctatcaaaa ttgtactaaa aggtattttg caatgtgtga ttggttaaca 5760
gattattttg ttttcaggat gcttgcacaa gcatggtcgg aaattttctc gaatatgtta 5820
ttgaaaatca ccgaatctac gaagagccag ccactacttt tcaggcattc cagatagccc 5880
tggtatgaca gcatttactt tgataattta tgcattgttt ccttcatcat ctgcctttgt 5940
ttagaatgtc ctcagaaggc agcactcctt tagttttaac tttccaatca taggattcaa 6000
atatccatta actggccttt gatcgctgca taatatatga atagttgaca tactgaatac 6060
gttgttaata atgcattttc agagttggat tgcagccttg ttggtaaagc aaattcttag 6120
ccacaaagaa tctctggtcc gtgcaaattc tgaattagct ttcaaatgct ctagagtaga 6180
ggtggattat atttattcga tattgtcctg catgaagagt ctgttcctgg agcatacaca 6240
aggtttgcag ttcgattgct ttggtactaa ttctaaacag tcagtggtta gcacaaaact 6300
agtaaatgaa agtctctcag ggctacagt gcgtgacgaa aagattaata cgaagtcgat 6360
gcgaaatagc tcagaggatg aagagtgcat gactgagaag agatgtagcc attatagcac 6420
agcaacaaga gatatcgaaa agactattag tggcataaaa aagaaataca agaagcaagt 6480
gcaaaagctt gtacaagagc atgaggaaaa gaaaatggag ctgttaaata tgtatgcaga 6540
caagaagcag aaacttgaaa ctagtaaaag tgtggaagca gcagtaattc gtattacctg 6600
ttcacggacc agtactcaag tgggtgatct caaactgctg gatcataatt atgaaagaaa 6660
gtttgatgaa atcaaaagtg agaaaaatga atgcctcaaa agtctggagc aaatgcacga 6720
ggttgcaaag aagaagttgg ctgaggatga agcctgttgg attaatcgga taaagagctg 6780
ggcagctaaa ttaaaagttt gtgttcccat tcaaagtggc aataacaagc attttagtgg 6840
ttcatcaaac atttcccaaa atgctcctga tgtacaaatt tgcaataatg ctaacgttga 6900
agctacttac gctgatacga attgcatggc ttccaaggtt aatcaagtgc cagaagcaga 6960
aaacacatta ggaaccatgt cgggtggcag cactcaacaa gttcatgaaa tggtggatgt 7020
aagaaatgac gagacaatgg atgtctcagc tttgtctcgt gaacagctta caaagagcca 7080
gtccaatgag cacgcttcta tcactgtgcc tgagattttg attcctgctg actgtcaaga 7140
ggaatttgcg gccttgaacg tgcatttgtc agaagaccag aattgtgaca gaataacatc 7200
tgcggcatca gatgaagatg tttcatcaag ggtgccagag gtatcccagt cactcgaaaa 7260
tctttctgcc tcccccgagt tttctctaaa tagagaggag gctttggtta caacagaaaa 7320
tagaagaaca agtcatgtgg gttttgatac tgataacatt ttggaccagc agaatagaga 7380
agattgttct cttgaccaag agattcctga cgagttagcg atgcctgtgc aacatcttgc 7440
gtctgtggta gagactaggg gtgctgctga atctgatcag gtacttactg gccctgtaga 7500
atagttgatg ccttgttcat ttaatctttt ctaatgttca ttcttgcttt cttgaaaata 7560
acgggtagtg atcagatgtc tttttttctc ttattaaatt cacttttctg gacagtatgg 7620
tcaagatata tgtcctatgc cttcttcact ggctggaaag caacctgacc cagcagcaaa 7680
cactgagagc gaaaatcttg aagaagcaat tgagcctcag tctgctggtt cagaaacagt 7740
agagactact gattttgctg catcacatca ggtccctatt gaagactttc ctttttttact 7800
agtttaaagt tatcaatctg tgttatgttc attctaagtt tccgtgagaa aaaggtgggg 7860
aaatgtggtt actgatcaag tctcgttgtt gttttaaatc gactcttttg acagggtgat 7920
caagttacat gtcctttgct atcttcaccg actggaaatc agcctgcgcc agaagcaaat 7980
attgaaggcc aaaatatcaa cacatcagct gagccccatg tagcgggtcc agatgcagta 8040
```

gagagtggtg attatgcagt aatagatcag gttattgcct taactaaaga caaatgtctt 8100
ttgttgttta aaagtcttac atctttgtaa tgctcgttct ggatatcctg caggaaacaa 8160
tgggtgctca ggatgcatgc tctctgccat ctggatcggt tggaactcag tctgacctag 8220
gagcaaacat tgagggtcaa aatgtcacaa cagtggctca acttcccaca gatggatcag 8280
atgcagttgt aaccggtgga tctcctgtat cagatcaggt acctgcctct gctcaaggac 8340
tttcttatgt gttggtttaa aggtctagtc cttagtaatg ttgaaactaa gcaaacagtg 8400
gatagtgatc atatggttat ttttgcttgt gaatttaata tttctggaca gtgtgcccag 8460
gatgcatctc ctatgccatt atcttcgcct ggaaatcacc ctgatacagc agttaatatc 8520
gagggtttag ataacacatc agtagctgag cctcatataa gtggatcaga tgcatgtgaa 8580
atggaaattt cagaacctgg tccccaagta gagcggtcaa cctttgcaag tcagtaactg 8640
ccttgggcat ttttaagtat cacctaggtc gacatatgtg attgccaaac agctaacaag 8700
gagatgcctt ttgtgcagat cttttccatg aaggtggcgt ggagcattca gcaggtgtaa 8760
cagctcttgt tccatcactt cttaacaatg gtacggaaca gattgccgtt caacctgttc 8820
ctcaaatacc tttccctgtg ttcaacgacc cgtttctgca tgaactggag aagttgcgga 8880
gagaatcaga gaactcaaag aagacttttg aagaaaaagt cagtttccct cattacccag 8940
ttacctcttg ttttggttta ttttctagct gcccattgac tctcagttgc ttgtgagcag 9000
aaatcaatct tgaaagctga actcgagagg aagatggctg aagtacaagc agagtttcga 9060
agaaaatttc atgaggtaga agccgagcat aacaccagaa cgacaaagat agagaaggat 9120
aagaatcttg ttataatgaa caaactgttg gcgaatgcgt tcttgtccaa atgtactgac 9180
aagaaggtat ctccctcagg agctccaagg ggtaagtgtc gaataatata gcaaattggt 9240
tttaaaaata aggcgacgaa gtcataatag cactttttct ccaggtaaaa ttcagcagct 9300
agcacagaga gcagcacaag tgagtgcact gagaaattac attgctcctc agcagcttca 9360
ggcatcttct tttcctgctc ctgctctggt ttcggctcct ctgcaacttc agcaatcatc 9420
atttcctgct cctggtccgg ctcctctgca gcctcaggca tcttcgtttc cttcttcagt 9480
ctctcgtcca tcagcccttc ttctgaattt tgcggtctgt ccaatgcctc agcccagaca 9540
gcctctcata tccaacatag ctccaactcc atcagttact cctgcaacaa atccaggtct 9600
gcgttctcct gcaccacacc taaactcata tagaccatcc tcttcaactc ccgtcgccac 9660
agctactcca acctcgtcag tgcctcctca agctttgaca tattcagctg tgtcaattca 9720
gcagcagcaa gaacaacaac cgcaacagag cttgagcagt ggattgcaga gcaacaatga 9780
agtggtttgt ctttctgacg acgagtgacc taagaggaga gatggttagg gtcttagtta 9840
ttgattttta gagagttaat aatagtatat atatatatgt ataagtaggt tacctaatct 9900
ctgtcgttaa tctaatttag tgagtcagga accgactcgt tggctaaggt ctctcctttt 9960
gaaacgcaac gttctacttt catgtatata aatacagtct gatcacacaa cacaaattga 10020
tgattgaaaa tactactgat ttaactttat agaaaaccca aattatagag cgacaacttt 10080
ataaacatgt caaacttcga agttaaaatt taagacccca taattttaca attatagatt 10140
ttaatactcc aactattttg tgatgttaaa agaagtatcc gagtcttttc tttccagttt 10200
ccccaccgtc ccatgactcc cccagccagt agaaaaagcc aaaaaagtaa acaaaaagtc 10260
gttaaaaaag ttaaattaaa aaaaaaatag atagttgacg tttactaaag tgatttgaat 10320
tgaacaatt 10329

<210> 2
<211> 6571
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (310)..(6312)

<400> 2
cacaagcatg agtttttcct tccggtaatc gtaaaatcaa aaatcgctcg ctcgagaaac 60

gccggtgcta gaagaggaaa gtaccgtaca taatcctgcg aacccaattc tcgtcttctt 120

caaactcagt tttccgaaac cccaaacacc gcgaggattg catggcctga agaaccactt 180

aatcgagaat tgtgctggaa ttctcaaatt ttccctcgcg tttttctttc acactctcgg 240

aatcggaaat ttccaccaag ctccgtcaag cgatagattc tgacaattac acactttcgc 300

gcagtgaat atg aag aaa gat gaa aag att ggt ttg acg ggg aga acc att 351
Met Lys Lys Asp Glu Lys Ile Gly Leu Thr Gly Arg Thr Ile
1 5 10

tac acc aga tcc cta gca gct tca att cct gcc tca gtt gaa caa gaa 399
Tyr Thr Arg Ser Leu Ala Ala Ser Ile Pro Ala Ser Val Glu Gln Glu
15 20 25 30

acc cct ggt ttg agg agg tca agc cgg ggg aca cca tct acg aag gta 447
Thr Pro Gly Leu Arg Arg Ser Ser Arg Gly Thr Pro Ser Thr Lys Val
35 40 45

ata act cca gct tct gct act aga aag tca gag aga ctg gct ccc tca 495
Ile Thr Pro Ala Ser Ala Thr Arg Lys Ser Glu Arg Leu Ala Pro Ser
50 55 60

cct gct tca gtt tca aaa aag tcc ggt gga atc gtc aag aat tcc aca 543
Pro Ala Ser Val Ser Lys Lys Ser Gly Gly Ile Val Lys Asn Ser Thr
65 70 75

cca agt tct ttg cga agg tcc aat agg ggg aag act gaa gta tcc ttg 591
Pro Ser Ser Leu Arg Arg Ser Asn Arg Gly Lys Thr Glu Val Ser Leu
80 85 90

cag agt tcc aaa gga tca gat aat tct atc agg aaa gga gat act tca 639
Gln Ser Ser Lys Gly Ser Asp Asn Ser Ile Arg Lys Gly Asp Thr Ser
95 100 105 110

ccg gat att gag cag aga aag gat agt gtt gaa gag tcg aca gat aag 687
Pro Asp Ile Glu Gln Arg Lys Asp Ser Val Glu Glu Ser Thr Asp Lys
115 120 125

atc aag cct ata atg tca gcc cga agt tac agg gca ttg ttt aga ggg 735
Ile Lys Pro Ile Met Ser Ala Arg Ser Tyr Arg Ala Leu Phe Arg Gly
130 135 140

aag ctc aag gaa tct gag gca tta gtt gat gct tcc cca aat gaa gag 783
Lys Leu Lys Glu Ser Glu Ala Leu Val Asp Ala Ser Pro Asn Glu Glu
145 150 155

gaa cta gta gtt gtt ggt tgt tct cgc cgc ata cct gca ggc aat gat 831
Glu Leu Val Val Val Gly Cys Ser Arg Arg Ile Pro Ala Gly Asn Asp
160 165 170

gat gtt caa ggt aaa aca gat tgt cca cca cct gca gat gca gga tca 879
Asp Val Gln Gly Lys Thr Asp Cys Pro Pro Pro Ala Asp Ala Gly Ser
175 180 185 190

aaa agg ctg cca gtt gac gaa act agt ttg gac aag ggc act gat ttt 927

```
Lys Arg Leu Pro Val Asp Glu Thr Ser Leu Asp Lys Gly Thr Asp Phe
                195                 200                 205

cct ttg aaa tca gtt acg gag acc gag aag ata gtg ctt gat gca tcc     975
Pro Leu Lys Ser Val Thr Glu Thr Glu Lys Ile Val Leu Asp Ala Ser
            210                 215                 220

ccc ata gtt gaa act ggg gat gac agt gtt ata ggt tca cca tct gag    1023
Pro Ile Val Glu Thr Gly Asp Asp Ser Val Ile Gly Ser Pro Ser Glu
        225                 230                 235

aat tta gag aca caa aag ctt caa gat ggt aag aca gat tgt tca cca    1071
Asn Leu Glu Thr Gln Lys Leu Gln Asp Gly Lys Thr Asp Cys Ser Pro
    240                 245                 250

cct gca aat gca gaa tcg aaa acg ctg cca gtt ggt gaa act agt tta    1119
Pro Ala Asn Ala Glu Ser Lys Thr Leu Pro Val Gly Glu Thr Ser Leu
255                 260                 265                 270

gaa aaa gaa tat cca caa aag ttt caa gat gat aac aca gat tgt cta    1167
Glu Lys Glu Tyr Pro Gln Lys Phe Gln Asp Asp Asn Thr Asp Cys Leu
                275                 280                 285

cca cct gca aat gca gaa tca aaa agg ctg cca gtt ggc gaa act agt    1215
Pro Pro Ala Asn Ala Glu Ser Lys Arg Leu Pro Val Gly Glu Thr Ser
            290                 295                 300

tta gaa aag gac act gat ttt cct ttg aaa tca act acg gag act gga    1263
Leu Glu Lys Asp Thr Asp Phe Pro Leu Lys Ser Thr Thr Glu Thr Gly
        305                 310                 315

aag atg gtt ctt tat gca tcc ccc ata gtt gaa act agg gat gac agc    1311
Lys Met Val Leu Tyr Ala Ser Pro Ile Val Glu Thr Arg Asp Asp Ser
    320                 325                 330

gtt ata tgt tca cca tct aca aat tta gaa acc caa aag ctt ctt gtc    1359
Val Ile Cys Ser Pro Ser Thr Asn Leu Glu Thr Gln Lys Leu Leu Val
335                 340                 345                 350

agt aaa act ggc tta gaa acc gac ata gtt ttg cct ttg aaa aga aaa    1407
Ser Lys Thr Gly Leu Glu Thr Asp Ile Val Leu Pro Leu Lys Arg Lys
                355                 360                 365

aga gac act gca gaa att gag ctg gat gca tgt gct aca gtt gca aat    1455
Arg Asp Thr Ala Glu Ile Glu Leu Asp Ala Cys Ala Thr Val Ala Asn
            370                 375                 380

gga gat gat cac gtt atg agt tct gat ggg gtc att cca tct cca tct    1503
Gly Asp Asp His Val Met Ser Ser Asp Gly Val Ile Pro Ser Pro Ser
        385                 390                 395

ggg tgc aaa aat gat aat cga cct gaa atg tgc aac acg tgt aaa aaa    1551
Gly Cys Lys Asn Asp Asn Arg Pro Glu Met Cys Asn Thr Cys Lys Lys
    400                 405                 410
```

```
cgg caa aag gtc aac ggt gat tgt caa aat agg agt gtt tgc tcc tgc    1599
Arg Gln Lys Val Asn Gly Asp Cys Gln Asn Arg Ser Val Cys Ser Cys
415                 420                 425                 430

att gtc cag cca gtt gaa gaa tct gat aac gtg aca cag gat atg aaa    1647
Ile Val Gln Pro Val Glu Glu Ser Asp Asn Val Thr Gln Asp Met Lys
                435                 440                 445

gaa act gga cca gtt acg agc aga gaa tat gag gag aac ggg caa ata    1695
Glu Thr Gly Pro Val Thr Ser Arg Glu Tyr Glu Glu Asn Gly Gln Ile
            450                 455                 460

caa cat ggt aaa tca agt gat ccc aaa ttc tat tct tcg gtg tac cca    1743
Gln His Gly Lys Ser Ser Asp Pro Lys Phe Tyr Ser Ser Val Tyr Pro
        465                 470                 475

gag tat tgg gtt cct gtg cag cta tca gat gta cag ctg gag caa tac    1791
Glu Tyr Trp Val Pro Val Gln Leu Ser Asp Val Gln Leu Glu Gln Tyr
    480                 485                 490

tgt cag act ctc ttc tcc aaa tcc tta tct ctt tct tca ctt tcg aag    1839
Cys Gln Thr Leu Phe Ser Lys Ser Leu Ser Leu Ser Ser Leu Ser Lys
495                 500                 505                 510

att gat ctt gga gct cta gaa gaa act ctc aat tct gta aga aaa acc    1887
Ile Asp Leu Gly Ala Leu Glu Glu Thr Leu Asn Ser Val Arg Lys Thr
                515                 520                 525

tgt gac cat cca tac gtt atg gat gca tct ttg aaa caa ctg ctc acc    1935
Cys Asp His Pro Tyr Val Met Asp Ala Ser Leu Lys Gln Leu Leu Thr
            530                 535                 540

aag aat ctg gag ttg cat gaa atc ctg gat gta gaa att aaa gcg agc    1983
Lys Asn Leu Glu Leu His Glu Ile Leu Asp Val Glu Ile Lys Ala Ser
        545                 550                 555

ggg aaa ctt cac ctc ctt gat aaa atg ctt act cat ata aaa aag aat    2031
Gly Lys Leu His Leu Leu Asp Lys Met Leu Thr His Ile Lys Lys Asn
    560                 565                 570

ggt tta aaa gca gtt gtc ttc tac cag gca aca caa acc cct gaa ggg    2079
Gly Leu Lys Ala Val Val Phe Tyr Gln Ala Thr Gln Thr Pro Glu Gly
575                 580                 585                 590

ctt ctg ctt ggt aat att ctc gaa gat ttt gtg ggt caa aga ttt ggt    2127
Leu Leu Leu Gly Asn Ile Leu Glu Asp Phe Val Gly Gln Arg Phe Gly
                595                 600                 605

cca aaa tct tat gag cat ggg ata tat tcc tca aag aag aac tcc gct    2175
Pro Lys Ser Tyr Glu His Gly Ile Tyr Ser Ser Lys Lys Asn Ser Ala
            610                 615                 620

ata aac aat ttc aac aag gag agt caa tgc tgt gtt ctg ctg ttg gaa    2223
Ile Asn Asn Phe Asn Lys Glu Ser Gln Cys Cys Val Leu Leu Leu Glu
        625                 630                 635
```

```
aca cgt gcc tgc agt caa acc att aaa ctc ttg cga gct gat gcg ttt    2271
Thr Arg Ala Cys Ser Gln Thr Ile Lys Leu Leu Arg Ala Asp Ala Phe
    640                 645                 650

att ctt ttt gga agc agc ttg aat cca tcg cat gat gtt aag cac gta    2319
Ile Leu Phe Gly Ser Ser Leu Asn Pro Ser His Asp Val Lys His Val
655                 660                 665                 670

gag aag ata aaa atc gag tca tgt tct gaa aga act aag ata ttc cga    2367
Glu Lys Ile Lys Ile Glu Ser Cys Ser Glu Arg Thr Lys Ile Phe Arg
                675                 680                 685

ttg tac tca gta tgt aca gtt gaa gaa aaa gcc ctg att ctg gct agg    2415
Leu Tyr Ser Val Cys Thr Val Glu Glu Lys Ala Leu Ile Leu Ala Arg
            690                 695                 700

caa aat atg cgg caa aat aaa gct gta gag aac cta aac cgc tct ctc    2463
Gln Asn Met Arg Gln Asn Lys Ala Val Glu Asn Leu Asn Arg Ser Leu
        705                 710                 715

acg cac gca ctg ctc atg tgg ggg gcg tca tac tta ttt gat aaa ctg    2511
Thr His Ala Leu Leu Met Trp Gly Ala Ser Tyr Leu Phe Asp Lys Leu
    720                 725                 730

gat cat ttt cac agc agt gaa act cca gat tca gga gtt tca ttt gaa    2559
Asp His Phe His Ser Ser Glu Thr Pro Asp Ser Gly Val Ser Phe Glu
735                 740                 745                 750

caa tct att atg gac ggc gtg att cat gaa ttc tcg tcc ata ctt tct    2607
Gln Ser Ile Met Asp Gly Val Ile His Glu Phe Ser Ser Ile Leu Ser
                755                 760                 765

tcc aaa ggt gga gaa gaa aat gaa gtc aag ctg tgt cta ctt ttg gag    2655
Ser Lys Gly Gly Glu Glu Asn Glu Val Lys Leu Cys Leu Leu Leu Glu
            770                 775                 780

gcc aag cat gct cag gga act tac agc agt gat tct act cta ttt ggt    2703
Ala Lys His Ala Gln Gly Thr Tyr Ser Ser Asp Ser Thr Leu Phe Gly
        785                 790                 795

gaa gac cat att aag ttg tca gat gaa gag agt cca aat ata ttt tgg    2751
Glu Asp His Ile Lys Leu Ser Asp Glu Glu Ser Pro Asn Ile Phe Trp
    800                 805                 810

tca aag ctg ttg ggg gga aaa aat cct atg tgg aaa tac cct tca gat    2799
Ser Lys Leu Leu Gly Gly Lys Asn Pro Met Trp Lys Tyr Pro Ser Asp
815                 820                 825                 830

act ccc caa agg aat cga aaa cga gtt cag tat ttt gag ggt tct gaa    2847
Thr Pro Gln Arg Asn Arg Lys Arg Val Gln Tyr Phe Glu Gly Ser Glu
                835                 840                 845

gcg agt ccc aaa act ggc gat ggt gga aat gca aag aag cga aag aag    2895
Ala Ser Pro Lys Thr Gly Asp Gly Gly Asn Ala Lys Lys Arg Lys Lys
```

850 855 860

gct tct gat gat gtc act gat ccc cgg gtc act gat ccg cca gta gat 2943
Ala Ser Asp Asp Val Thr Asp Pro Arg Val Thr Asp Pro Pro Val Asp
865 870 875

gat gat gaa aga aag gcc tct ggg aag gat cac atg ggg gct ttg gag 2991
Asp Asp Glu Arg Lys Ala Ser Gly Lys Asp His Met Gly Ala Leu Glu
880 885 890

tca cca aaa gtc ata aca ctc cag tca tca tgt aaa tct tct ggt aca 3039
Ser Pro Lys Val Ile Thr Leu Gln Ser Ser Cys Lys Ser Ser Gly Thr
895 900 905 910

gat ggt aca ttg gat gga aat gat gct ttt ggc ttg tat tct atg ggc 3087
Asp Gly Thr Leu Asp Gly Asn Asp Ala Phe Gly Leu Tyr Ser Met Gly
915 920 925

agc cat atc tct gga atc cca gag gat atg tta gct agt caa gat tgg 3135
Ser His Ile Ser Gly Ile Pro Glu Asp Met Leu Ala Ser Gln Asp Trp
930 935 940

ggg aaa ata ccg gat gaa tca cag agg agg ctc cac act gtt tta aag 3183
Gly Lys Ile Pro Asp Glu Ser Gln Arg Arg Leu His Thr Val Leu Lys
945 950 955

ccg aag atg gca aaa ctt tgc caa gtt ttg cat ctt tca gat gct tgc 3231
Pro Lys Met Ala Lys Leu Cys Gln Val Leu His Leu Ser Asp Ala Cys
960 965 970

aca agc atg gtc gga aat ttt ctc gaa tat gtt att gaa aat cac cga 3279
Thr Ser Met Val Gly Asn Phe Leu Glu Tyr Val Ile Glu Asn His Arg
975 980 985 990

atc tac gaa gag cca gcc act act ttt cag gca ttc cag ata gcc ctg 3327
Ile Tyr Glu Glu Pro Ala Thr Thr Phe Gln Ala Phe Gln Ile Ala Leu
995 1000 1005

agt tgg att gca gcc ttg ttg gta aag caa att ctt agc cac aaa gaa 3375
Ser Trp Ile Ala Ala Leu Leu Val Lys Gln Ile Leu Ser His Lys Glu
1010 1015 1020

tct ctg gtc cgt gca aat tct gaa tta gct ttc aaa tgc tct aga gta 3423
Ser Leu Val Arg Ala Asn Ser Glu Leu Ala Phe Lys Cys Ser Arg Val
1025 1030 1035

gag gtg gat tat att tat tcg ata ttg tcc tgc atg aag agt ctg ttc 3471
Glu Val Asp Tyr Ile Tyr Ser Ile Leu Ser Cys Met Lys Ser Leu Phe
1040 1045 1050

ctg gag cat aca caa ggt ttg cag ttc gat tgc ttt ggt act aat tct 3519
Leu Glu His Thr Gln Gly Leu Gln Phe Asp Cys Phe Gly Thr Asn Ser
1055 1060 1065 1070

aaa cag tca gtg gtt agc aca aaa cta gta aat gaa agt ctc tca ggg 3567

Lys Gln Ser Val Val Ser Thr Lys Leu Val Asn Glu Ser Leu Ser Gly
1075 1080 1085

gct aca gtg cgt gac gaa aag att aat acg aag tcg atg cga aat agc 3615
Ala Thr Val Arg Asp Glu Lys Ile Asn Thr Lys Ser Met Arg Asn Ser
1090 1095 1100

tca gag gat gaa gag tgc atg act gag aag aga tgt agc cat tat agc 3663
Ser Glu Asp Glu Glu Cys Met Thr Glu Lys Arg Cys Ser His Tyr Ser
1105 1110 1115

aca gca aca aga gat atc gaa aag act att agt ggc ata aaa aag aaa 3711
Thr Ala Thr Arg Asp Ile Glu Lys Thr Ile Ser Gly Ile Lys Lys Lys
1120 1125 1130

tac aag aag caa gtg caa aag ctt gta caa gag cat gag gaa aag aaa 3759
Tyr Lys Lys Gln Val Gln Lys Leu Val Gln Glu His Glu Glu Lys Lys
1135 1140 1145 1150

atg gag ctg tta aat atg tat gca gac aag aag cag aaa ctt gaa act 3807
Met Glu Leu Leu Asn Met Tyr Ala Asp Lys Lys Gln Lys Leu Glu Thr
1155 1160 1165

agt aaa agt gtg gaa gca gca gta att cgt att acc tgt tca cgg acc 3855
Ser Lys Ser Val Glu Ala Ala Val Ile Arg Ile Thr Cys Ser Arg Thr
1170 1175 1180

agt act caa gtg ggt gat ctc aaa ctg ctg gat cat aat tat gaa aga 3903
Ser Thr Gln Val Gly Asp Leu Lys Leu Leu Asp His Asn Tyr Glu Arg
1185 1190 1195

aag ttt gat gaa atc aaa agt gag aaa aat gaa tgc ctc aaa agt ctg 3951
Lys Phe Asp Glu Ile Lys Ser Glu Lys Asn Glu Cys Leu Lys Ser Leu
1200 1205 1210

gag caa atg cac gag gtt gca aag aag aag ttg gct gag gat gaa gcc 3999
Glu Gln Met His Glu Val Ala Lys Lys Lys Leu Ala Glu Asp Glu Ala
1215 1220 1225 1230

tgt tgg att aat cgg ata aag agc tgg gca gct aaa tta aaa gtt tgt 4047
Cys Trp Ile Asn Arg Ile Lys Ser Trp Ala Ala Lys Leu Lys Val Cys
1235 1240 1245

gtt ccc att caa agt ggc aat aac aag cat ttt agt ggt tca tca aac 4095
Val Pro Ile Gln Ser Gly Asn Asn Lys His Phe Ser Gly Ser Ser Asn
1250 1255 1260

att tcc caa aat gct cct gat gta caa att tgc aat aat gct aac gtt 4143
Ile Ser Gln Asn Ala Pro Asp Val Gln Ile Cys Asn Asn Ala Asn Val
1265 1270 1275

gaa gct act tac gct gat acg aat tgc atg gct tcc aag gtt aat caa 4191
Glu Ala Thr Tyr Ala Asp Thr Asn Cys Met Ala Ser Lys Val Asn Gln
1280 1285 1290

gtg cca gaa gca gaa aac aca tta gga acc atg tcg ggt ggc agc act 4239
Val Pro Glu Ala Glu Asn Thr Leu Gly Thr Met Ser Gly Gly Ser Thr
1295 1300 1305 1310

caa caa gtt cat gaa atg gtg gat gta aga aat gac gag aca atg gat 4287
Gln Gln Val His Glu Met Val Asp Val Arg Asn Asp Glu Thr Met Asp
1315 1320 1325

gtc tca gct ttg tct cgt gaa cag ctt aca aag agc cag tcc aat gag 4335
Val Ser Ala Leu Ser Arg Glu Gln Leu Thr Lys Ser Gln Ser Asn Glu
1330 1335 1340

cac gct tct atc act gtg cct gag att ttg att cct gct gac tgt caa 4383
His Ala Ser Ile Thr Val Pro Glu Ile Leu Ile Pro Ala Asp Cys Gln
1345 1350 1355

gag gaa ttt gcg gcc ttg aac gtg cat ttg tca gaa gac cag aat tgt 4431
Glu Glu Phe Ala Ala Leu Asn Val His Leu Ser Glu Asp Gln Asn Cys
1360 1365 1370

gac aga ata aca tct gcg gca tca gat gaa gat gtt tca tca agg gtg 4479
Asp Arg Ile Thr Ser Ala Ala Ser Asp Glu Asp Val Ser Ser Arg Val
1375 1380 1385 1390

cca gag gta tcc cag tca ctc gaa aat ctt tct gcc tcc ccc gag ttt 4527
Pro Glu Val Ser Gln Ser Leu Glu Asn Leu Ser Ala Ser Pro Glu Phe
1395 1400 1405

tct cta aat aga gag gag gct ttg gtt aca aca gaa aat aga aga aca 4575
Ser Leu Asn Arg Glu Glu Ala Leu Val Thr Thr Glu Asn Arg Arg Thr
1410 1415 1420

agt cat gtg ggt ttt gat act gat aac att ttg gac cag cag aat aga 4623
Ser His Val Gly Phe Asp Thr Asp Asn Ile Leu Asp Gln Gln Asn Arg
1425 1430 1435

gaa gat tgt tct ctt gac caa gag att cct gac gag tta gcg atg cct 4671
Glu Asp Cys Ser Leu Asp Gln Glu Ile Pro Asp Glu Leu Ala Met Pro
1440 1445 1450

gtg caa cat ctt gcg tct gtg gta gag act agg ggt gct gct gaa tct 4719
Val Gln His Leu Ala Ser Val Val Glu Thr Arg Gly Ala Ala Glu Ser
1455 1460 1465 1470

gat cag tat ggt caa gat ata tgt cct atg cct tct tca ctg gct gga 4767
Asp Gln Tyr Gly Gln Asp Ile Cys Pro Met Pro Ser Ser Leu Ala Gly
1475 1480 1485

aag caa cct gac cca gca gca aac act gag agc gaa aat ctt gaa gaa 4815
Lys Gln Pro Asp Pro Ala Ala Asn Thr Glu Ser Glu Asn Leu Glu Glu
1490 1495 1500

gca att gag cct cag tct gct ggt tca gaa aca gta gag act act gat 4863
Ala Ile Glu Pro Gln Ser Ala Gly Ser Glu Thr Val Glu Thr Thr Asp
1505 1510 1515

ttt gct gca tca cat cag ggt gat caa gtt aca tgt cct ttg cta tct 4911
Phe Ala Ala Ser His Gln Gly Asp Gln Val Thr Cys Pro Leu Leu Ser
1520 1525 1530

tca ccg act gga aat cag cct gcg cca gaa gca aat att gaa ggc caa 4959
Ser Pro Thr Gly Asn Gln Pro Ala Pro Glu Ala Asn Ile Glu Gly Gln
1535 1540 1545 1550

aat atc aac aca tca gct gag ccc cat gta gcg ggt cca gat gca gta 5007
Asn Ile Asn Thr Ser Ala Glu Pro His Val Ala Gly Pro Asp Ala Val
1555 1560 1565

gag agt ggt gat tat gca gta ata gat cag gaa aca atg ggt gct cag 5055
Glu Ser Gly Asp Tyr Ala Val Ile Asp Gln Glu Thr Met Gly Ala Gln
1570 1575 1580

gat gca tgc tct ctg cca tct gga tcg gtt gga act cag tct gac cta 5103
Asp Ala Cys Ser Leu Pro Ser Gly Ser Val Gly Thr Gln Ser Asp Leu
1585 1590 1595

gga gca aac att gag ggt caa aat gtc aca aca gtg gct caa ctt ccc 5151
Gly Ala Asn Ile Glu Gly Gln Asn Val Thr Thr Val Ala Gln Leu Pro
1600 1605 1610

aca gat gga tca gat gca gtt gta acc ggt gga tct cct gta tca gat 5199
Thr Asp Gly Ser Asp Ala Val Val Thr Gly Gly Ser Pro Val Ser Asp
1615 1620 1625 1630

cag tgt gcc cag gat gca tct cct atg cca tta tct tcg cct gga aat 5247
Gln Cys Ala Gln Asp Ala Ser Pro Met Pro Leu Ser Ser Pro Gly Asn
1635 1640 1645

cac cct gat aca gca gtt aat atc gag ggt tta gat aac aca tca gta 5295
His Pro Asp Thr Ala Val Asn Ile Glu Gly Leu Asp Asn Thr Ser Val
1650 1655 1660

gct gag cct cat ata agt gga tca gat gca tgt gaa atg gaa att tca 5343
Ala Glu Pro His Ile Ser Gly Ser Asp Ala Cys Glu Met Glu Ile Ser
1665 1670 1675

gaa cct ggt ccc caa gta gag cgg tca acc ttt gca aat ctt ttc cat 5391
Glu Pro Gly Pro Gln Val Glu Arg Ser Thr Phe Ala Asn Leu Phe His
1680 1685 1690

gaa ggt ggc gtg gag cat tca gca ggt gta aca gct ctt gtt cca tca 5439
Glu Gly Gly Val Glu His Ser Ala Gly Val Thr Ala Leu Val Pro Ser
1695 1700 1705 1710

ctt ctt aac aat ggt acg gaa cag att gcc gtt caa cct gtt cct caa 5487
Leu Leu Asn Asn Gly Thr Glu Gln Ile Ala Val Gln Pro Val Pro Gln
1715 1720 1725

ata cct ttc cct gtg ttc aac gac ccg ttt ctg cat gaa ctg gag aag 5535
Ile Pro Phe Pro Val Phe Asn Asp Pro Phe Leu His Glu Leu Glu Lys

```
        1730                1735                1740

ttg cgg aga gaa tca gag aac tca aag aag act ttt gaa gaa aaa aaa     5583
Leu Arg Arg Glu Ser Glu Asn Ser Lys Lys Thr Phe Glu Glu Lys Lys
    1745                1750                1755

tca atc ttg aaa gct gaa ctc gag agg aag atg gct gaa gta caa gca     5631
Ser Ile Leu Lys Ala Glu Leu Glu Arg Lys Met Ala Glu Val Gln Ala
1760                1765                1770

gag ttt cga aga aaa ttt cat gag gta gaa gcc gag cat aac acc aga     5679
Glu Phe Arg Arg Lys Phe His Glu Val Glu Ala Glu His Asn Thr Arg
1775                1780                1785                1790

acg aca aag ata gag aag gat aag aat ctt gtt ata atg aac aaa ctg     5727
Thr Thr Lys Ile Glu Lys Asp Lys Asn Leu Val Ile Met Asn Lys Leu
                1795                1800                1805

ttg gcg aat gcg ttc ttg tcc aaa tgt act gac aag aag gta tct ccc     5775
Leu Ala Asn Ala Phe Leu Ser Lys Cys Thr Asp Lys Lys Val Ser Pro
            1810                1815                1820

tca gga gct cca agg ggt aaa att cag cag cta gca cag aga gca gca     5823
Ser Gly Ala Pro Arg Gly Lys Ile Gln Gln Leu Ala Gln Arg Ala Ala
        1825                1830                1835

caa gtg agt gca ctg aga aat tac att gct cct cag cag ctt cag gca     5871
Gln Val Ser Ala Leu Arg Asn Tyr Ile Ala Pro Gln Gln Leu Gln Ala
    1840                1845                1850

tct tct ttt cct gct cct gct ctg gtt tcg gct cct ctg caa ctt cag     5919
Ser Ser Phe Pro Ala Pro Ala Leu Val Ser Ala Pro Leu Gln Leu Gln
1855                1860                1865                1870

caa tca tca ttt cct gct cct ggt ccg gct cct ctg cag cct cag gca     5967
Gln Ser Ser Phe Pro Ala Pro Gly Pro Ala Pro Leu Gln Pro Gln Ala
                1875                1880                1885

tct tcg ttt cct tct tca gtc tct cgt cca tca gcc ctt ctt ctg aat     6015
Ser Ser Phe Pro Ser Ser Val Ser Arg Pro Ser Ala Leu Leu Leu Asn
            1890                1895                1900

ttt gcg gtc tgt cca atg cct cag ccc aga cag cct ctc ata tcc aac     6063
Phe Ala Val Cys Pro Met Pro Gln Pro Arg Gln Pro Leu Ile Ser Asn
        1905                1910                1915

ata gct cca act cca tca gtt act cct gca aca aat cca ggt ctg cgt     6111
Ile Ala Pro Thr Pro Ser Val Thr Pro Ala Thr Asn Pro Gly Leu Arg
    1920                1925                1930

tct cct gca cca cac cta aac tca tat aga cca tcc tct tca act ccc     6159
Ser Pro Ala Pro His Leu Asn Ser Tyr Arg Pro Ser Ser Ser Thr Pro
1935                1940                1945                1950

gtc gcc aca gct act cca acc tcg tca gtg cct cct caa gct ttg aca     6207
```

```
Val Ala Thr Ala Thr Pro Thr Ser Ser Val Pro Pro Gln Ala Leu Thr
               1955                1960                1965

tat tca gct gtg tca att cag cag cag caa gaa caa caa ccg caa cag     6255
Tyr Ser Ala Val Ser Ile Gln Gln Gln Gln Glu Gln Gln Pro Gln Gln
          1970                1975                1980

agc ttg agc agt gga ttg cag agc aac aat gaa gtg gtt tgt ctt tct     6303
Ser Leu Ser Ser Gly Leu Gln Ser Asn Asn Glu Val Val Cys Leu Ser
      1985                1990                1995

gac gac gag tgacctaaga ggagagatgg ttagggtctt agttattgat             6352
Asp Asp Glu
  2000

ttttagagag ttaataatag tatatatata tatgtataag taggttacct aatctctgtc 6412

gttaatctaa tttagtgagt caggaaccga ctcgttggct aaggtctctc cttttgaaac 6472

gcaacgttct actttcatgt atataaatac agtctgatca cacaacacaa attgatgatt 6532

gaaaatacta ctgatttaac ttaaaaaaaa aaaaaaaaa                          6571


<210> 3
<211> 2001
<212> PRT
<213> Arabidopsis thaliana

<400> 3
Met Lys Lys Asp Glu Lys Ile Gly Leu Thr Gly Arg Thr Ile Tyr Thr
  1               5                  10                  15

Arg Ser Leu Ala Ala Ser Ile Pro Ala Ser Val Glu Gln Glu Thr Pro
            20                  25                  30

Gly Leu Arg Arg Ser Ser Arg Gly Thr Pro Ser Thr Lys Val Ile Thr
        35                  40                  45

Pro Ala Ser Ala Thr Arg Lys Ser Glu Arg Leu Ala Pro Ser Pro Ala
    50                  55                  60

Ser Val Ser Lys Lys Ser Gly Gly Ile Val Lys Asn Ser Thr Pro Ser
65                  70                  75                  80

Ser Leu Arg Arg Ser Asn Arg Gly Lys Thr Glu Val Ser Leu Gln Ser
                85                  90                  95

Ser Lys Gly Ser Asp Asn Ser Ile Arg Lys Gly Asp Thr Ser Pro Asp
            100                 105                 110

Ile Glu Gln Arg Lys Asp Ser Val Glu Glu Ser Thr Asp Lys Ile Lys
        115                 120                 125

Pro Ile Met Ser Ala Arg Ser Tyr Arg Ala Leu Phe Arg Gly Lys Leu
```

130 135 140

Lys Glu Ser Glu Ala Leu Val Asp Ala Ser Pro Asn Glu Glu Glu Leu
145 150 155 160

Val Val Val Gly Cys Ser Arg Arg Ile Pro Ala Gly Asn Asp Asp Val
165 170 175

Gln Gly Lys Thr Asp Cys Pro Pro Pro Ala Asp Ala Gly Ser Lys Arg
180 185 190

Leu Pro Val Asp Glu Thr Ser Leu Asp Lys Gly Thr Asp Phe Pro Leu
195 200 205

Lys Ser Val Thr Glu Thr Glu Lys Ile Val Leu Asp Ala Ser Pro Ile
210 215 220

Val Glu Thr Gly Asp Asp Ser Val Ile Gly Ser Pro Ser Glu Asn Leu
225 230 235 240

Glu Thr Gln Lys Leu Gln Asp Gly Lys Thr Asp Cys Ser Pro Pro Ala
245 250 255

Asn Ala Glu Ser Lys Thr Leu Pro Val Gly Glu Thr Ser Leu Glu Lys
260 265 270

Glu Tyr Pro Gln Lys Phe Gln Asp Asp Asn Thr Asp Cys Leu Pro Pro
275 280 285

Ala Asn Ala Glu Ser Lys Arg Leu Pro Val Gly Glu Thr Ser Leu Glu
290 295 300

Lys Asp Thr Asp Phe Pro Leu Lys Ser Thr Thr Glu Thr Gly Lys Met
305 310 315 320

Val Leu Tyr Ala Ser Pro Ile Val Glu Thr Arg Asp Asp Ser Val Ile
325 330 335

Cys Ser Pro Ser Thr Asn Leu Glu Thr Gln Lys Leu Leu Val Ser Lys
340 345 350

Thr Gly Leu Glu Thr Asp Ile Val Leu Pro Leu Lys Arg Lys Arg Asp
355 360 365

Thr Ala Glu Ile Glu Leu Asp Ala Cys Ala Thr Val Ala Asn Gly Asp
370 375 380

Asp His Val Met Ser Ser Asp Gly Val Ile Pro Ser Pro Ser Gly Cys
385 390 395 400

Lys Asn Asp Asn Arg Pro Glu Met Cys Asn Thr Cys Lys Lys Arg Gln
405 410 415

Lys Val Asn Gly Asp Cys Gln Asn Arg Ser Val Cys Ser Cys Ile Val
420 425 430

Gln Pro Val Glu Glu Ser Asp Asn Val Thr Gln Asp Met Lys Glu Thr
435 440 445

Gly Pro Val Thr Ser Arg Glu Tyr Glu Glu Asn Gly Gln Ile Gln His
450 455 460

Gly Lys Ser Ser Asp Pro Lys Phe Tyr Ser Ser Val Tyr Pro Glu Tyr
465 470 475 480

Trp Val Pro Val Gln Leu Ser Asp Val Gln Leu Glu Gln Tyr Cys Gln
485 490 495

Thr Leu Phe Ser Lys Ser Leu Ser Leu Ser Ser Leu Ser Lys Ile Asp
500 505 510

Leu Gly Ala Leu Glu Glu Thr Leu Asn Ser Val Arg Lys Thr Cys Asp
515 520 525

His Pro Tyr Val Met Asp Ala Ser Leu Lys Gln Leu Leu Thr Lys Asn
530 535 540

Leu Glu Leu His Glu Ile Leu Asp Val Glu Ile Lys Ala Ser Gly Lys
545 550 555 560

Leu His Leu Leu Asp Lys Met Leu Thr His Ile Lys Lys Asn Gly Leu
565 570 575

Lys Ala Val Val Phe Tyr Gln Ala Thr Gln Thr Pro Glu Gly Leu Leu
580 585 590

Leu Gly Asn Ile Leu Glu Asp Phe Val Gly Gln Arg Phe Gly Pro Lys
595 600 605

Ser Tyr Glu His Gly Ile Tyr Ser Ser Lys Lys Asn Ser Ala Ile Asn
610 615 620

Asn Phe Asn Lys Glu Ser Gln Cys Cys Val Leu Leu Leu Glu Thr Arg
625 630 635 640

Ala Cys Ser Gln Thr Ile Lys Leu Leu Arg Ala Asp Ala Phe Ile Leu
645 650 655

Phe Gly Ser Ser Leu Asn Pro Ser His Asp Val Lys His Val Glu Lys
660 665 670

Ile Lys Ile Glu Ser Cys Ser Glu Arg Thr Lys Ile Phe Arg Leu Tyr
675 680 685

Ser Val Cys Thr Val Glu Glu Lys Ala Leu Ile Leu Ala Arg Gln Asn
690 695 700

Met Arg Gln Asn Lys Ala Val Glu Asn Leu Asn Arg Ser Leu Thr His
705 710 715 720

Ala Leu Leu Met Trp Gly Ala Ser Tyr Leu Phe Asp Lys Leu Asp His
725 730 735

Phe His Ser Ser Glu Thr Pro Asp Ser Gly Val Ser Phe Glu Gln Ser
740 745 750

Ile Met Asp Gly Val Ile His Glu Phe Ser Ser Ile Leu Ser Ser Lys
755 760 765

Gly Gly Glu Glu Asn Glu Val Lys Leu Cys Leu Leu Leu Glu Ala Lys
770 775 780

His Ala Gln Gly Thr Tyr Ser Ser Asp Ser Thr Leu Phe Gly Glu Asp
785 790 795 800

His Ile Lys Leu Ser Asp Glu Glu Ser Pro Asn Ile Phe Trp Ser Lys
805 810 815

Leu Leu Gly Gly Lys Asn Pro Met Trp Lys Tyr Pro Ser Asp Thr Pro
820 825 830

Gln Arg Asn Arg Lys Arg Val Gln Tyr Phe Glu Gly Ser Glu Ala Ser
835 840 845

Pro Lys Thr Gly Asp Gly Gly Asn Ala Lys Lys Arg Lys Lys Ala Ser
850 855 860

Asp Asp Val Thr Asp Pro Arg Val Thr Asp Pro Pro Val Asp Asp Asp
865 870 875 880

Glu Arg Lys Ala Ser Gly Lys Asp His Met Gly Ala Leu Glu Ser Pro
885 890 895

Lys Val Ile Thr Leu Gln Ser Ser Cys Lys Ser Ser Gly Thr Asp Gly
900 905 910

Thr Leu Asp Gly Asn Asp Ala Phe Gly Leu Tyr Ser Met Gly Ser His
915 920 925

Ile Ser Gly Ile Pro Glu Asp Met Leu Ala Ser Gln Asp Trp Gly Lys
930 935 940

Ile Pro Asp Glu Ser Gln Arg Arg Leu His Thr Val Leu Lys Pro Lys
945 950 955 960

Met Ala Lys Leu Cys Gln Val Leu His Leu Ser Asp Ala Cys Thr Ser
965 970 975

Met Val Gly Asn Phe Leu Glu Tyr Val Ile Glu Asn His Arg Ile Tyr
980 985 990

Glu Glu Pro Ala Thr Thr Phe Gln Ala Phe Gln Ile Ala Leu Ser Trp
995 1000 1005

Ile Ala Ala Leu Leu Val Lys Gln Ile Leu Ser His Lys Glu Ser Leu

1010 1015 1020

Val Arg Ala Asn Ser Glu Leu Ala Phe Lys Cys Ser Arg Val Glu Val
025 1030 1035 1040

Asp Tyr Ile Tyr Ser Ile Leu Ser Cys Met Lys Ser Leu Phe Leu Glu
1045 1050 1055

His Thr Gln Gly Leu Gln Phe Asp Cys Phe Gly Thr Asn Ser Lys Gln
1060 1065 1070

Ser Val Val Ser Thr Lys Leu Val Asn Glu Ser Leu Ser Gly Ala Thr
1075 1080 1085

Val Arg Asp Glu Lys Ile Asn Thr Lys Ser Met Arg Asn Ser Ser Glu
1090 1095 1100

Asp Glu Glu Cys Met Thr Glu Lys Arg Cys Ser His Tyr Ser Thr Ala
105 1110 1115 1120

Thr Arg Asp Ile Glu Lys Thr Ile Ser Gly Ile Lys Lys Lys Tyr Lys
1125 1130 1135

Lys Gln Val Gln Lys Leu Val Gln Glu His Glu Glu Lys Lys Met Glu
1140 1145 1150

Leu Leu Asn Met Tyr Ala Asp Lys Lys Gln Lys Leu Glu Thr Ser Lys
1155 1160 1165

Ser Val Glu Ala Ala Val Ile Arg Ile Thr Cys Ser Arg Thr Ser Thr
1170 1175 1180

Gln Val Gly Asp Leu Lys Leu Leu Asp His Asn Tyr Glu Arg Lys Phe
185 1190 1195 1200

Asp Glu Ile Lys Ser Glu Lys Asn Glu Cys Leu Lys Ser Leu Glu Gln
1205 1210 1215

Met His Glu Val Ala Lys Lys Lys Leu Ala Glu Asp Glu Ala Cys Trp
1220 1225 1230

Ile Asn Arg Ile Lys Ser Trp Ala Ala Lys Leu Lys Val Cys Val Pro
1235 1240 1245

Ile Gln Ser Gly Asn Asn Lys His Phe Ser Gly Ser Ser Asn Ile Ser
1250 1255 1260

Gln Asn Ala Pro Asp Val Gln Ile Cys Asn Asn Ala Asn Val Glu Ala
265 1270 1275 1280

Thr Tyr Ala Asp Thr Asn Cys Met Ala Ser Lys Val Asn Gln Val Pro
1285 1290 1295

Glu Ala Glu Asn Thr Leu Gly Thr Met Ser Gly Gly Ser Thr Gln Gln
1300 1305 1310

Val His Glu Met Val Asp Val Arg Asn Asp Glu Thr Met Asp Val Ser
1315 1320 1325

Ala Leu Ser Arg Glu Gln Leu Thr Lys Ser Gln Ser Asn Glu His Ala
1330 1335 1340

Ser Ile Thr Val Pro Glu Ile Leu Ile Pro Ala Asp Cys Gln Glu Glu
345 1350 1355 1360

Phe Ala Ala Leu Asn Val His Leu Ser Glu Asp Gln Asn Cys Asp Arg
1365 1370 1375

Ile Thr Ser Ala Ala Ser Asp Glu Asp Val Ser Ser Arg Val Pro Glu
1380 1385 1390

Val Ser Gln Ser Leu Glu Asn Leu Ser Ala Ser Pro Glu Phe Ser Leu
1395 1400 1405

Asn Arg Glu Glu Ala Leu Val Thr Thr Glu Asn Arg Arg Thr Ser His
1410 1415 1420

Val Gly Phe Asp Thr Asp Asn Ile Leu Asp Gln Gln Asn Arg Glu Asp
425 1430 1435 1440

Cys Ser Leu Asp Gln Glu Ile Pro Asp Glu Leu Ala Met Pro Val Gln
1445 1450 1455

His Leu Ala Ser Val Val Glu Thr Arg Gly Ala Ala Glu Ser Asp Gln
1460 1465 1470

Tyr Gly Gln Asp Ile Cys Pro Met Pro Ser Ser Leu Ala Gly Lys Gln
1475 1480 1485

Pro Asp Pro Ala Ala Asn Thr Glu Ser Glu Asn Leu Glu Glu Ala Ile
1490 1495 1500

Glu Pro Gln Ser Ala Gly Ser Glu Thr Val Glu Thr Thr Asp Phe Ala
505 1510 1515 1520

Ala Ser His Gln Gly Asp Gln Val Thr Cys Pro Leu Leu Ser Ser Pro
1525 1530 1535

Thr Gly Asn Gln Pro Ala Pro Glu Ala Asn Ile Glu Gly Gln Asn Ile
1540 1545 1550

Asn Thr Ser Ala Glu Pro His Val Ala Gly Pro Asp Ala Val Glu Ser
1555 1560 1565

Gly Asp Tyr Ala Val Ile Asp Gln Glu Thr Met Gly Ala Gln Asp Ala
1570 1575 1580

Cys Ser Leu Pro Ser Gly Ser Val Gly Thr Gln Ser Asp Leu Gly Ala
585 1590 1595 1600

Asn Ile Glu Gly Gln Asn Val Thr Thr Val Ala Gln Leu Pro Thr Asp
1605 1610 1615

Gly Ser Asp Ala Val Val Thr Gly Gly Ser Pro Val Ser Asp Gln Cys
1620 1625 1630

Ala Gln Asp Ala Ser Pro Met Pro Leu Ser Ser Pro Gly Asn His Pro
1635 1640 1645

Asp Thr Ala Val Asn Ile Glu Gly Leu Asp Asn Thr Ser Val Ala Glu
1650 1655 1660

Pro His Ile Ser Gly Ser Asp Ala Cys Glu Met Glu Ile Ser Glu Pro
665 1670 1675 1680

Gly Pro Gln Val Glu Arg Ser Thr Phe Ala Asn Leu Phe His Glu Gly
1685 1690 1695

Gly Val Glu His Ser Ala Gly Val Thr Ala Leu Val Pro Ser Leu Leu
1700 1705 1710

Asn Asn Gly Thr Glu Gln Ile Ala Val Gln Pro Val Pro Gln Ile Pro
1715 1720 1725

Phe Pro Val Phe Asn Asp Pro Phe Leu His Glu Leu Glu Lys Leu Arg
1730 1735 1740

Arg Glu Ser Glu Asn Ser Lys Lys Thr Phe Glu Glu Lys Lys Ser Ile
745 1750 1755 1760

Leu Lys Ala Glu Leu Glu Arg Lys Met Ala Glu Val Gln Ala Glu Phe
1765 1770 1775

Arg Arg Lys Phe His Glu Val Glu Ala Glu His Asn Thr Arg Thr Thr
1780 1785 1790

Lys Ile Glu Lys Asp Lys Asn Leu Val Ile Met Asn Lys Leu Leu Ala
1795 1800 1805

Asn Ala Phe Leu Ser Lys Cys Thr Asp Lys Lys Val Ser Pro Ser Gly
1810 1815 1820

Ala Pro Arg Gly Lys Ile Gln Gln Leu Ala Gln Arg Ala Ala Gln Val
825 1830 1835 1840

Ser Ala Leu Arg Asn Tyr Ile Ala Pro Gln Gln Leu Gln Ala Ser Ser
1845 1850 1855

Phe Pro Ala Pro Ala Leu Val Ser Ala Pro Leu Gln Leu Gln Gln Ser
1860 1865 1870

Ser Phe Pro Ala Pro Gly Pro Ala Pro Leu Gln Pro Gln Ala Ser Ser
1875 1880 1885

Phe Pro Ser Ser Val Ser Arg Pro Ser Ala Leu Leu Leu Asn Phe Ala

1890 1895 1900

Val Cys Pro Met Pro Gln Pro Arg Gln Pro Leu Ile Ser Asn Ile Ala
905 1910 1915 1920

Pro Thr Pro Ser Val Thr Pro Ala Thr Asn Pro Gly Leu Arg Ser Pro
1925 1930 1935

Ala Pro His Leu Asn Ser Tyr Arg Pro Ser Ser Ser Thr Pro Val Ala
1940 1945 1950

Thr Ala Thr Pro Thr Ser Ser Val Pro Pro Gln Ala Leu Thr Tyr Ser
1955 1960 1965

Ala Val Ser Ile Gln Gln Gln Gln Glu Gln Gln Pro Gln Gln Ser Leu
1970 1975 1980

Ser Ser Gly Leu Gln Ser Asn Asn Glu Val Val Cys Leu Ser Asp Asp
985 1990 1995 2000

Glu

<210> 4
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Oligonucleotide

<400> 4
catctacggc aatgtaccag c 21

<210> 5
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Oligonucleotide

<400> 5
gatgggaatt ggctgagtgg c 21

<210> 6
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 6
cagttccaaa cgtaaaacgg c 21

<210> 7
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 7
ntcgastwts gwgtt 15

<210> 8
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 8
ngtcgaswga nawgaa 16

<210> 9
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 9
wgtgnagwan canaga 16

<210> 10
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic

Oligonucleotide

<400> 10
wggwancwga wangca 16

<210> 11
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 11
wcgwwgawca ngncga 16

<210> 12
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 12
wgcnagtnag wanaag 16

<210> 13
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 13
awgcangncw ganata 16

<210> 14
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 14

ctgtacatac tgagtacaat cgga 24

<210> 15
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 15
gcttcaattc ctgcctcagt tgaac 25

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 16
ctctacgtgc ttaacatcat gcga 24

<210> 17
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 17
ccagcttctg ctactagaaa gtcag 25

<210> 18
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 18
ctggagttgc atgaaatcct ggatg 25

<210> 19
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 19
gctctttgta agctgttcac gagac 25

<210> 20
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 20
tcgcatgatg ttaagcacgt agag 24

<210> 21
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 21
gagtactggt ccgtgaacag gtaat 25

<210> 22
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 22
atgcttgcac aagcatggtc ggaaa 25

<210> 23
<211> 25
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 23
tgcaacatcg tgcatttgct ccaga 25

<210> 24
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 24
cacaagcatg agtttttcct tccgg 25

<210> 25
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 25
ctgactttct agtagcagaa gctgg 25

<210> 26
<211> 519
<212> DNA
<213> Brassica oleracea

<220>
<223> seq1-23

<400> 26
gaattcctgn nttacggcat ccttaataga ctgttcaaat ggaactcctg aacctggggt 60
tccactccca tggaagtgtt ccagcttatc aaataaatat gatgcccccc acatgagcaa 120
tgcatgtgtg agaggacggt ttaggttctc tagaggctta ttttgcctag caagaatcag 180
ggttttttct tcaactgtaa acactgagta caaccggaaa atcttagttc tttcagaaca 240
cgactcaacc tttatcttct ctaagagctt aacgtcatgc gatggattca ggctgcttcc 300
aaaaagtata aaagactcag cgcgtaagag tttaatgctt tgactacagg cacgtatttc 360
cagcagcaga ataaaacact cactctcctt gttgaaattg tttatagcgt tcttcttcga 420
gaggcagacc ccatgctcat aggaattttg accaaatctt tgcatcagaa aatcttcgag 480
aatattacca agcagaagcc cctcagggct atgtattgc 519

<210> 27
<211> 419
<212> DNA
<213> Brassica oleracea

<220>
<223> seq1-27

<400> 27
gaattcagga tcaaaagggt tgccggttgg agaaactggt ttagagaaag gctctgattt 60
tcctgtggaa gtaactaagg atatagagaa gacagtggtt gattcatccc ccatggttga 120
aactgaggat ggcagtgtta taggttcacc atccgagaat ccagaaccac aaaagcttcg 180
tgacagtgaa actagcttgg aaaccgatat agacttggct ctgaaaagaa aaagagacac 240
tgcagaaatt gtgatggatg catgtacaaa tgcagatgac cgcattatga gtactgatgg 300
ggttattcct tttccacccg tgtgcacaaa tattaatcaa cccgaaaggt gtggcacatg 360
tcaaaaacgg caaaagtaag aatttccgac tgttgtctgt cgttttgaaa ccatttgcc 419

<210> 28
<211> 467
<212> DNA
<213> Brassica oleracea

<220>
<223> seq1-43

<400> 28
gaattctcgt ccatactttc ttccgatgtt ggagaagaaa atgaaggcaa gctgtgtcta 60
cttttggaag ccaagcatgc tcagggaagt tacagcactg atgctactct atttggtgaa 120
gaacatgtca agttatcaga tgaaagtcca aatatgtttt ggtcaaagct gttgagtgga 180
aagaacccta tgtggaaata ctgttcggat actcctcaaa ggagtcgaaa aagagtacgg 240
catcttcagg gctatgagga gactaccaaa gttggcaatg gcggaaactt aaagaagaaa 300
aagaaggctt cagatgatgt cacagtagat aacgctgaga gaaaagcctc tggaaaggat 360
cacatgggta aaacagttca cttcctgctc ctttacctct agtgttcatt gaatgttcca 420
tttactttgc ttactatctt tccttcaggg catttggagt caccaaa 467

<210> 29
<211> 490
<212> DNA
<213> Brassica oleracea

<220>
<223> seq1-47

<400> 29
gaattcagct tttaaaactg atctctgctc acagataatt taagagtcag tgaaaattga 60
gataaaacga accaaaactg gaggtaacag atactctgag aacaactaac cttttcttca 120
taagtcttct ttgtgttctc tgattctctc cgcagcttct ccagttcatg ctgaaatggg 180
tcactgaaca cagggaaagg tacttgagga acaggtggag tggcattctg tcccgtagca 240
ttgttaagct gtgaagaaac aggagctgtt acacctgctg gaggctccac aacaccttca 300
tcgacaacgt ctgcgtaaaa ggtattacca gattgtcagt ttctctggca aacacatacg 360
ttatacttaa atgcaaaaga gcagttactg acttgcaaag gttggttgtt ctacttgagc 420

atcaggttct gctacttcca tttcacatgc ttctgatcca gttgtgcgag gcgcagccat 480
tgttgtgttg 490

<210> 30
<211> 515
<212> DNA
<213> Brassica oleracea

<220>
<223> 2-33

<400> 30
tctagagaag aggtggatta tgtatattct tttctgtact gcatgaagag tctattcgtg 60
gggcgcacac aaggtttcca agaaaagggt gaagaatgca tggctgagaa aagaggtagc 120
cattatagct cagtaaccaa ggatgttgaa aagactatta gcgacatcaa aaagaaatgc 180
agtaagagcc tgcataagct tgtacaaacc ctcgaggaag aaaagatgga cctgatgaat 240
aggaatgctg tcaagaagca ggaacttcag aattgtaaaa aggtggaagc atcatttatt 300
cgtgtcacct attcaggtat aaatactcag agcttacatg atgctctcca acggctggaa 360
tgtacttttg aaagaaagtt tgatgatctc aaaggagagt tggatgaatg ccttgaaagt 420
ttagagcaaa taaacgaggc tggaaagaag aagttggctg aagatgaagc ctgttggatt 480
agtcggatag agaaatgggc acgagctgaa ttaag 515

<210> 31
<211> 574
<212> DNA
<213> Brassica oleracea

<220>
<223> seq2-37

<400> 31
tctagaccaa actattaaac gctaaacata agaagattag atcactcgtc atcagagaga 60
cagaccacat cattgctcct ctgcaatcca ctccccaagt tctgtggttg ttcttgctgc 120
tgaataaacg catttgaata tggtaaaggg ttggagatga gaggttgtct tggttgaggc 180
attgtgcagt acggagccga agcagtatga ttcctcagtg cgcttacttg tgttgctctc 240
tgtgctagct gctggattct aactggagaa agaaaaaaag aaaaaaaagg tgttattatg 300
acttcataac cttatatctt taaaaaacaa ttatgcttct attattcgaa cacttgccca 360
ttggagttgc tgctgaggaa tgagaggaga ttctgctcgt acatttagac aagaacgcac 420
tcgacaacag cttgttcttt ataacaagat tcttcctcgt ctgtaacttc gtctttctgg 480
ctgcatgtac agcttgtacc tcatgaaact ttctctgata ctcttcttgt aattcagcta 540
tcttcttctc gaatttagct ttcaagactg cttt 574

<210> 32
<211> 466
<212> DNA
<213> Brassica oleracea

<220>
<223> seq2-53

<400> 32
tctagattgt aattttaaat ttacaacaaa ttttgaaagg gtcagcgatg agtttgcaaa 60

```
tctccgtgtt tcctccagca ttgctcagcc agttcaagaa cctgatcact tggcacaggt 120
tggtttcttc ttgctttact ttggacacct gtttaatatt ggcctgtcaa atttacttat 180
ccttttactt ctaaactgca aattctggtc tgcattgcat tgtgatatga aggtatctgg 240
acccgcttca agcagagact atggggagga caggcagaat atgcaacaag ataaatcaca 300
tgaccgaaag ttgtcatcga tgtatccaga gtattgggtt ccagtgcagc tatcagatgt 360
acagatagag caatactgtc ggactctctt ctccaaatct tcatctcttt cttcgctgtc 420
gaggactgat cctgttcgag ctcttgaaca aactctcagt tctgta                466
```

```
<210> 33
<211> 417
<212> DNA
<213> Brassica oleracea

<220>
<223> seq2-57

<400> 33
tctagagcaa ttgaaaccta attccgattt tgcgcgggcc agagattctt cacggttgaa 60
cttttgctta acgaaagaga ctgcaatcca aatctggaag tgcattatta agaacgtatt 120
cagcaatatt cataaattat gcaacaatca aaggccttac gttgtggcct acaaagcatg 180
gattttgtta gatattagta gctagtctaa ttcaagcaat taatggaagt ttctatccta 240
tgactggaaa gttaaacatt cccacaaaag cagtgatgcc acagatgatg aagaagaaaa 300
atgcatatac tatggaagtg aatgctatca taccacagct atctggaagg cctgcaatgt 360
tgtagctggc tctttgcaga cacggtggtt gtcaataata tattcaagaa ctttttc    417
```

Gene involved in epigenetic gene silencing

The present invention relates to DNA which encodes proteins that control gene silencing, and particularly the silencing of plant genes.

The loss of expression of previously active genes in plants, also referred to as gene silencing, is observed in response to developmental, environmental or unknown signals. It occurs at a frequency higher than that of mutations, yet it is markedly stable during somatic transmission. Gene silencing, initially perceived as an unwanted source of instability of transgene expression, is now regarded as a molecular tool to intentionally regulate gene expression.

It appears that chromosomal position or structure of the affected loci are factors determining the frequency and strength of silencing. Inactivation seems to preferentially affect genes present in multiple copies and is thought to be a consequence of sequence redundancy. Many examples of homology-dependent gene silencing have been reported. Closer analysis has allowed the classification of silencing events according to the relative position of the affected loci (*cis*, *trans*, allelic, ectopic), the origin of the affected genes (endogenous or transgenic), and the level of interaction (transcriptional or post-transcriptional). While post-transcriptional silencing seems to mainly involve the formation of aberrant RNA molecules and is occasionally, but not necessarily, accompanied by DNA methylation, silencing interfering with transcription initiation is more strictly correlated with hypermethylation of the DNA and possibly with alteration of chromatin structure at the silent loci. It is, however, not clear whether these molecular events are a prerequisite for gene silencing or a consequence of the silent state.

In the case of transcriptional silencing, the inactive state of silenced genes is stably transmitted through mitotic and meiotic divisions. As in other organisms, trans-acting modifier loci are assumed to be responsible for the stability of the inactive state of the silenced genes. Mutations in such loci resulting in mutated proteins are expected to result in reduced gene silencing and reactivation of previously silent loci by interfering with the maintenance of the silent state, or by a failure to recognize sequence redundancy. It has been reported that mutations in the DDM1 gene of Arabidopsis thaliana release

transcriptional gene silencing and that this genes encodes a SWI2/SNF2-like protein involved in chromatin remodeling. However, mutation of the DDM1 gene causes severe pleiotropic effects. Therefore, to be able to modify such effects making use of gene technology, it is necessary to identify further specific modifier loci and characterize the corresponding wild-type and mutant proteins. It is the main objective of the present invention to provide DNA comprising an open reading frame encoding such a protein.

Trans-acting modifier loci according to the present invention can be identified by T-DNA insertion mutagenesis as described in Example 1 for an Arabidopsis line carrying a heritably inactivated, methylated hygromycin resistance gene. A mutation of a silencing modifier locus results in release of silencing of the hygromycin resistance gene and restores hygromycin resistance. Plants homozygous for the silent resistance gene are subjected to transformation with a selectable marker gene different from the hygromycin resistance gene, which is under the control of the T-DNA 1'-2' dual promoter. Transformants are selected and their progeny screened for hygromycin resistance. The mutant phenotype (hygromycin resistance) is screened for genetic co-segregation with a specific T-DNA insert. Cloning of the tagged gene using routine methods of recombinant DNA technology allows to characterize the mutant and wild-type DNA sequence of the silencing modifier locus as well as the encoded protein.

Within the context of the present invention reference to a gene is to be understood as reference to a DNA coding sequence associated with regulatory sequences, which allow transcription of the coding sequence into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Examples of regulatory sequences are promoter sequences, 5' and 3' untranslated sequences, introns, and termination sequences.

A promoter is understood to be a DNA sequence initiating transcription of an associated DNA sequence, and may also include elements that act as regulators of gene expression such as activators, enhancers, or repressors.

Expression of a gene refers to its transcription into RNA or its transcription and subsequent translation into protein within a living cell. In the case of antisense constructs expression refers to the transcription of the antisense DNA only.

The term transformation of cells designates the introduction of nucleic acid into a host cell, particularly the stable integration of a DNA molecule into the genome of said cell.

Any part or piece of a specific nucleotide or amino acid sequence is referred to as a component sequence.

DNA according to the present invention comprises an open reading frame encoding a protein characterized by an amino acid sequence comprising a component sequence of at least 150 amino acid residues having 40% or more identity with SEQ ID NO: 3. In particular the protein encoded by the open reading frame can be described by the formula R_1-R_2-R_3, wherein

-- R_1, R_2 and R_3 constitute component sequences consisting of amino acid residues independently selected from the group of the amino acid residues Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg, and His,
-- R_1 and R_3 consist independently of 0 to 3000 amino acid residues;
-- R_2 consists of at least 150 amino acid residues; and
-- R_2 is at least 40% identical to an aligned component sequence of SEQ ID NO: 3.

In most cases the total length of the protein will be in the range of 1000 to 3000 amino acid residues. In preferred embodiments of the invention the component sequence R_2 consists of at least 200 amino acid residues. Specific examples of the component sequence R_2 are component sequences of SEQ ID NO: 3 represented by the following range of amino acids:

1 - 416 (corresponding to exon 2);
418 - 583 (corresponding to exons 3 to 5);
584 - 890 (corresponding to exon 6);
892 - 1472 (corresponding to exons 7 to 9);
1007 - 1472 (corresponding to exon 9);
1473 - 1631 (corresponding to exons 10 to 12);
1632 - 1827 (corresponding to exons 13 to 15); and
1829 - 2001 (corresponding to exon 16).

In a preferred embodiment of the present invention at least one of the component sequences R_1 or R_3 comprises one or more additional component sequences with a length of at least 50 amino acids and at least 60% identical to an aligned component sequence of SEQ ID NO: 3. Specific examples of such additional component sequences are component sequences of SEQ ID NO: 3 represented by the following range of amino acids:

420 - 525 (corresponding to exons 3 and 4);
444 - 525 (corresponding to exon 4);
526 - 583 (corresponding to exon 5);
892 - 971 (corresponding to exon 7);
892 - 1006 (corresponding to exons 7 and 8);
1473 - 1524 (corresponding to exon 10);
1525 - 1576 (corresponding to exon 11);
1577 - 1631 (corresponding to exon 12);
1632 - 1690 (corresponding to exons 13);
1692 - 1757 (corresponding to exons 14); and
1758 - 1827 (corresponding to exons 15).

Particularly preferred embodiments of the DNA according to the present invention encode a protein having a component sequence defined by amino acids 478-490, 584-600, 617-630, 654-668, 676-690, 718-734, 776-788, 1222-1233, 1738-1749 or 1761-1770 of SEQ ID NO: 3. Preferably, the encoded protein comprises at least two, three or more different representatives of said component sequences. Specific examples of said embodiments encode a protein characterized by the amino acid sequence of SEQ ID NO: 3, an allelic amino acid sequence having amino acid residue K instead of M at position 705 of SEQ ID NO: 3, or an amino acid residue D instead of E at position 1219 of SEQ ID NO: 3.

Dynamic programming algorithms yield different kinds of alignments. In general there exist two approaches towards sequence alignment. Algorithms as proposed by Needleman & Wunsch and by Sellers align the entire length of two sequences providing a global alignment of the sequences. The Smith-Waterman algorithm on the other hand yields local alignments. A local alignment aligns the pair of regions within the sequences that are most similar given the choice of scoring matrix and gap penalties. This allows a database search to focus on the most highly conserved regions of the sequences. It also allows similar domains within sequences to be identified. To speed up alignments using the Smith-Waterman algorithm both BLAST (Basic Local Alignment Search Tool) and FASTA place additional restrictions on the alignments.

Within the context of the present invention alignments are conveniently performed using BLAST, a set of similarity search programs designed to explore all of the available

sequence databases regardless of whether the query is protein or DNA. Version BLAST 2.0 (Gapped BLAST) of this search tool has been made publicly available on the internet (currently http://www.ncbi.nlm.nih.gov/BLAST/). It uses a heuristic algorithm which seeks local as opposed to global alignments and is therefore able to detect relationships among sequences which share only isolated regions. The scores assigned in a BLAST search have a well-defined statistical interpretation. Particularly useful within the scope of the present invention are the blastp program allowing for the introduction of gaps in the local sequence alignments and the PSI-BLAST program, both programs comparing an amino acid query sequence against a protein sequence database, as well as a blastp variant program allowing local alignment of two sequences only. Said programs are preferably run with optional parameters set to the default values.

Sequence alignments using BLAST can also take into account whether the substitution of one amino acid for another is likely to conserve the physical and chemical properties necessary to maintain the structure and function of the protein or is more likely to disrupt essential structural and functional features of a protein. Such sequence similarity is quantified in terms of a percentage of "positive" amino acids, as compared to the percentage of identical amino acids and can help assigning a protein to the correct protein family in border-line cases.

Sequence alignments using such computer programs reveal the presence of an ATP/GTP-binding motif A (amino acids 460 to 467 in SEQ ID NO:3), the consensus sequence of which is (Ala/Gly)XaaXaaXaaXaaGlyLys(Ser/Thr), wherein (Ala/Gly) indicates Ala or Gly, Xaa indicates any naturally occurring amino acid and (Ser/Thr) indicates Ser or Thr. Alignment additionally reveals a region (amino acid position 479 to 719 in SEQ ID: 3), similar to part of the ATPase/helicase domain of proteins in the SWI2/SNF2 family which are involved in chromatin remodeling but no significant overall sequence identity with known proteins.

Specific examples of DNA according to the present invention are described in SEQ ID NO: 1 and SEQ ID NO: 2 encoding an Arabidopsis protein described in SEQ ID NO: 3. Stretches of SEQ ID NO: 3 having 50 to 500 amino acids length can show between 20 and 50% sequence identity to stretches of known protein sequences after alignment. Overall alignments of SEQ ID NO: 3, however, result in sequence identities lower than 30%. Thus,

the present invention defines a new protein family the members of which are characterized by an amino acid sequence comprising a component sequence of at least 150 amino acid residues having 40% or more identity with an aligned component sequence of SEQ ID NO: 3. Preferably the amino acid sequence identity is higher than 50% or even higher than 55%.

DNA encoding proteins belonging to the new protein family according to the present invention can be isolated from monocotyledonous and dicotyledonous plants. Preferred sources are corn, sugarbeet, sunflower, winter oilseed rape, soybean, cotton, wheat, rice, potato, broccoli, cauliflower, cabbage, cucumber, sweet corn, daikon, garden beans, lettuce, melon, pepper, squash, tomato, or watermelon. However, they can also be isolated from mammalian sources such as mouse or human tissues. The following general method, can be used, which the person skilled in the art knows to adapt to the specific task. A single stranded fragment of SEQ ID NO: 1 or SEQ ID NO: 2 consisting of at least 15, preferably 20 to 30 or even more than 100 consecutive nucleotides is used as a probe to screen a DNA library for clones hybridizing to said fragment. The factors to be observed for hybridization are described in Sambrook et al, Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, chapters 9.47-9.57 and 11.45-11.49, 1989. Hybridizing clones are sequenced and DNA of clones comprising a complete coding region encoding a protein characterized by an amino acid sequence comprising a component sequence of at least 150 amino acid residues having 40% or more sequence identity to SEQ ID NO: 3 is purified. Said DNA can then be further processed by a number of routine recombinant DNA techniques such as restriction enzyme digestion, ligation, or polymerase chain reaction analysis.

The disclosure of SEQ ID NO: 1 and SEQ ID NO: 2 enables a person skilled in the art to design oligonucleotides for polymerase chain reactions which attempt to amplify DNA fragments from templates comprising a sequence of nucleotides characterized by any continuous sequence of 15 and preferably 20 to 30 or more basepairs in SEQ ID NO: 1 or SEQ ID NO: 2. Said nucleotides comprise a sequence of nucleotides which represents 15 and preferably 20 to 30 or more basepairs of SEQ ID NO: 1 or SEQ ID NO: 2. Polymerase chain reactions performed using at least one such oligonucleotide and their amplification products constitute another embodiment of the present invention.

EXAMPLES:

Example 1: ***T-DNA Insertion***

Transgenic line A of *Arabidopsis thaliana* ecotype Zürich with a transcriptionally silenced locus containing multiple copies of a chimeric hygromycin phosphotransferase gene (*hpt*) has been described in Mittelsten Scheid et al, Mol Gen Genet 228: 104-112, 1991 and Mittelsten Scheid et al, Proc Natl Acad Sci USA 93: 7114-7119, 1996. A homozygous, diploid genotype of said line is subjected to *Agrobacterium* mediated gene transfer by *in planta* vacuum infiltration (Bechtold et al., C R Acad Sci Paris Life Science 316: 1194-1199, 1993) generating more than 4000 independent T-DNA transformants. The binary vector with T-DNA consisting of the coding region of the *bar* gene transcriptionally fused to the 1' promoter (p1'barbi), the *Agrobacterium* strain (C58ClRif^R) and the transformation protocol are described by Mengiste et al, Plant J 12: 945-948, 1997. Transformants (T1 plants) are selected by repeated spraying of germinated seedlings with Basta solution (150 mg/l) and grown to maturity.

Example 2: ***Mutant Selection***

Selfed seeds (T2 families) are collected from individual transformants. Prior to screening for revertants of the silenced phenotype, seeds are dried for one week at room temperature and cold-treated at 4°C for a minimum of one week. Pooled aliquots of approximately 1000 seeds (consisting of 50 seeds from 20 T2 families) are surface-sterilized twice (with 5% sodium hypochlorite containing 0.1% Tween 80) for 7 min and washed with sterile double-distilled water. For selection, each aliquot is plated on 14-cm Petri dishes containing 75 ml germination medium (according to Masson et al, Plant J 2: 829-933, 1992) solidified with 0.8% agar and containing 10 mg/l hygromycin B (Calbiochem). To ensure equal distribution during sowing, seeds are mixed with 30 ml of the same medium containing 0.4% agar. As positive control two seeds from a hygromycin-resistant line are sown at marked locations on each plate. Plates are cold-treated at 4°C for 2 days and subsequently subjected to alternating periods of 16 hours light at 21°C and 8 hours darkness at 16°C. Hygromycin resistance is evaluated each day for 8-15 days after sowing.

Example 3: ***Molecular and Genetic Analysis of the Mutant***

Following identification of 11 hygromycin-resistant seedlings in one of the pools, the families forming this pool are re-screened individually. One family contains approximately 25% hygromycin-resistant seedlings. Six resistant plantlets of this family are transferred to larger vessels containing germination medium without hygromycin. After rosette formation and development of the root system, plants are transferred to soil for further growth and seed setting. Prior to potting, tissue explants are taken from each plant to generate callus cultures on RCA medium (Table 1) with or without 10 mg/l hygromycin B. Callus cultures are used as a source of material for DNA and RNA analyses and for a further confirmation of hygromycin resistance in this tissue.
Genomic DNA is isolated using a CTAB based method as described by Mittelsten Scheid et al, Mol Gen Genet 244: 325-330, 1994, and incubated with restriction enzymes *BamHI, HpaII, MspII, DraI, EcoRV, RcaI* or *HindIII.* Total RNA is obtained using a RNAeasy kit (Qiagen) according to the supplier's recommendation. Southern and northern blot analysis are performed under conditions described by Church and Gilbert, Proc Natl Acad Sci USA 81: 1991-1995, 1984, using DNA fragments labeled with ^{32}P by random prime labeling. The coding region of the *hpt* gene, or DNA consisting of the P35S promoter, *hpt* coding and terminator region, or the coding region of the *bar* gene together with the 1' promoter are used as probes.

Northern blot analysis of 4 hygromycin-resistant siblings shows restoration of transcription of the *hpt* gene. Southern blot analysis of said siblings indicates that there is no detectable rearrangement within the complex *hpt* insert. The *hpt* transgene complex in the mutant is still hypermethylated like in the original line A, as judged by Southern blot analysis with the methylation-sensitive restriction enzymes *HpaII* and *MspI,* and by genomic sequencing of the promoter region after treatment with bisulfate. There is also no influence of the mutation on the methylation of repetitive genomic DNA in contrast to that observed for the *som* mutations.

The hygromycin-resistant plants, as well as non-selected siblings from the same family are grown to set seeds, checked for Basta resistance in the next generation, and scored for the number and size of the T-DNA inserts by Southern analysis. The results demonstrate that the original T-DNA transformant must have contained 2 T-DNA insertions segregating

independently in the siblings. One insert co-segregates with the hygromycin resistant mutant phenotype. A plant homozygous for this insert and lacking the other T-DNA insert, is used for cloning the corresponding T-DNA insertion site.

Histochemical GUS staining of crosses between plants with mutant phenotype and the transgenic plant line GUS-TS (obtainable from Dr. H. Vaucheret, INRA, Versailles Cedex, France) of Arabidopsis thaliana ecotype Colombia containing a transcriptionally silenced locus with multiple copies of a chimeric beta-glucuronidase (gus) gene reveals reactivation of the silent GUS gene in the F2 progeny which are homozygous for the mom allele.

Inbreeding of plants with the mom1 mutant phenotype does not result in any morphological abnormalities even in the 9th generation of inbreeding. This is in contrast to the *som* mutants.

Backcrossing of the mutant phenotype of mom1 with line A (see example 1) results in immediate resilencing of the reactivated hpt gene upon introduction of a wild-type MOM allele in F1 hybrids. This also is in contrast to the *som* mutants.

Table 1: Composition of RCA medium

RCA medium	
MS macro 10 x	100 ml
B5 micro 1000 x	1 ml
ferric citrate	5 ml
NT vitamins 100 x	10 ml
sucrose	10 g
MES	5 ml
agar	10 g
NAA	0.1 mg
BAP	1 mg
pH 5.8 (KOH)	
ad 1 l	

MS macro 10 x	
potassium nitrate	19 g
ammonium nitrate	16.5 g
calcium chloride (x 2 H_2O)	4.4 g
magnesium sulfate (x 7 H_2O)	3.7 g
potassium dihydrogen phosphate	1.7 g
ad 1 l	

B5 micro 1000 x	
magnesium sulfate (x H_2O)	1000 mg
boric acid	300 mg
zinc sulfate (x 7 H_2O)	200 mg
potassium iodide	75 mg
sodium molybdate (x 2 H_2O)	25 mg
copper sulfate (x 5 H_2O)	2.5 mg
cobalt chloride (x 6 H_2O)	2.5 mg
ad 100 ml	

ferric citrate	
ammonium iron citrate	10 g
ad 1 l	

NT vitamins 100 x	
myo-inositol	1000 mg
thiamine HCl	10 mg
ad 1 l	

MES	
MES	14 g
pH 6 (NaOH)	
ad 100 ml	

Example 4: ***Cloning of the "Silencing Gene"***

Genomic DNA from the plant containing only the T-DNA co-segregating with the hygromycin resistant mutant phenotype is isolated. The DNA is subjected to TAIL (thermal asymmetric interlaced) PCR according to Liu et al, Plant J 8: 457-463, 1995, using 3 specific, nested primers close to the right border of the T-DNA (5'-CAT CTA CGG CAA TGT ACC AGC-3' (SEQ ID NO: 4), 5'-GAT GGG AAT TGG CTG AGT GGC-3' (SEQ ID NO: 5), 5'-CAG TTC CAA ACG TAA AAC GGC-3' (SEQ ID NO: 6)) which are directed outwards, and one of several degenerate primers which might bind in flanking plant DNA. Two out of the following seven degenerate primers

AD1	5'-NTC GAS TWT SGW GTT-3'	(Liu et al supra; SEQ ID NO: 7)
AD2	5'-NGT CGA SWG ANA WGA A-3'	(Liu et al supra; SEQ ID NO: 8)
AD3	5'-WGT GNA GWA NCA NAG A-3'	(Liu et al supra; SEQ ID NO: 9)
AD4	5'-WGG WAN CWG AWA NGC A-3'	(SEQ ID NO: 10)
AD5	5'-WCG WWG AWC ANG NCG A-3'	(SEQ ID NO: 11)
AD6	5'-WGC NAG TNA GWA NAA G-3'	(SEQ ID NO: 12)
AD7	5'-AWG CAN GNC WGA NAT A-3'	(SEQ ID NO: 13)

actually result in amplification of specific fragments. The larger one obtained using AD7 is cloned and sequenced. It contains 50 bp of the T-DNA and 275 bp of flanking plant DNA. In Southern blot analysis it is shown that this PCR fragment contains the plant DNA flanking the T-DNA. The PCR fragment is used to screen a genomic library (Stratagene) of wild type *Arabidopsis thaliana* ecotype Columbia. Three genomic clones hybridizing to the PCR fragment are identified. The genomic clones are further mapped with restriction enzymes, hybridized to the PCR fragment and aligned to each other. In one of the genomic clones obtained (p4A-11), the sequence found to flank the T-DNA of the insertion mutation is located approximately in the middle of the genomic sequence. An approximately 800 bp EcoRI-SalI fragment of p4A-11 is used to obtain the overlapping genomic clone p5-6, and an approximately 700 bp EcoRI fragment of p5-6 is used to obtain genomic clone p30-1 overlapping with p5-6. An approximately 700 bp HindIII fragment of p30-1 is used to obtain the genomic clone p33-19 overlapping with p30-1. Said clones are sequenced to design primers for RT-PCR. The approximately 700 bp EcoRI fragment of p5-6 is further used for screening of a cDNA library of wild type *Arabidopsis thaliana* ecotype Zurich according to

Elledge et al, Proc Natl Acad Sci USA 88: 1731-1735, 1991). Nine cDNA clones are obtained and the longest clone p17-8 having a length of 2.6 kb is sequenced.

Example 5: ***Sequence Analysis and Alignments***

Taking into account the large size of the Arabidopsis silencing gene cloned above it cannot be entirely excluded that the authentic nucleotide and amino acid sequences of the gene and protein, respectively, might deviate from the sequences given in SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 3 at a few positions due to mutations arising from the cloning procedure or due to ambiguities in the sequencing reactions. Additionally, sequencing of DNA derived from a different ecotype can reveal allelic differences. Thus, the sequences of SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 3 represent the corresponding genes and proteins of *Arabidopsis thaliana* ecotype Zurich, whereas genomic sequences obtained from *Arabidopsis thaliana* ecotype Columbia reveal two mismatches at nucleotide positions 4338 (A instead of T) and 6721 (T instead of G) of SEQ ID NO: 1, which result in an amino acid residue K instead of M at position 705 of SEQ ID NO: 3 and an amino acid residue D instead of E at position 1219 of SEQ ID NO: 3.

The 2.6 kb cDNA clone is analyzed sequentially from both ends and is shown to contain one large ORF as well as a 3' untranslated sequence.

Analysis of the genomic clones reveals that clones p4A-11 and p5-6 contain sequences homologous to the cDNA sequence as well as 7 intron sequences. Comparing the genomic sequences with the DNA sequences flanking the T-DNA insert, it turns out that the T-DNA insertion causes a deletion of about 2 kb of genomic DNA. The 5' end of the deletion is located in an intron (intron 12) and the 3' end of the deletion is located downstream of the 3' end of the cDNA. The sequence of 5' end of the cDNA clone terminates in the middle of the sequence of the genomic clone p5-6. Three independent nested RT-PCR reactions are performed to obtain additional cDNA sequences further upstream. The sequences of the primers used for these RT-PCRs are as follows:

RT1-1	5'-CTGTACATACTGAGTACAATCGGA-3'	(SEQ ID NO: 14)
RT1-2	5'-GCTTCAATTCCTGCCTCAGTTGAAC-3'	(SEQ ID NO: 15)
RT1-3	5'-CTCTACGTGCTTAACATCATGCGA-3'	(SEQ ID NO: 16)
RT1-4	5'-CCAGCTTCTGCTACTAGAAAGTCAG-3'	(SEQ ID NO: 17)

RT2/3-1	5'-CTGGAGTTGCATGAAATCCTGGATG-3'	(SEQ ID NO: 18)
RT2/3-2	5'-GCTCTTTGTAAGCTGTTCACGAGAC-3'	(SEQ ID NO: 19)
RT2-3	5'-TCGCATGATGTTAAGCACGTAGAG-3'	(SEQ ID NO: 20)
RT2-4	5'-GAGTACTGGTCCGTGAACAGGTAAT-3'	(SEQ ID NO: 21)
RT3-3	5'-ATGCTTGCACAAGCATGGTCGGAAA-3'	(SEQ ID NO: 22)
RT3-4	5'-TGCAACATCGTGCATTTGCTCCAGA-3'	(SEQ ID NO: 23)
RT4-1	5'-CACAAGCATGAGTTTTTCCTTCCGG-3'	(SEQ ID NO: 24)
RT4-2	5'-CTGACTTTCTAGTAGCAGAAGCTGG-3'	(SEQ ID NO: 25)

Sequences of several parts of the genomic clones are found to be deposited in the *Arabidopsis* database (accession numbers B67281, B62563, B20434, B20425, B21274, B08967, B11993, B20116, B12496 and B10852 as end sequences of BAC, and Z18494 and AA597930 as partial cDNA sequences, on 13 Apr 1999). A comparison of the encoded protein sequence with the Swiss Protein Database reveals partial similarity with ATPase/helicase proteins of the SWI2/SNF2 family (amino acid position 479 to 719 in SEQ ID NO: 3). The encoded protein consists of 2001 amino acids and is calculated to have a molecular weight of 219 kD and a pI of 5.1. An ATP/GTP-binding motif (amino acid position 460 to 467 in SEQ ID NO: 3) and three nuclear localization motifs (amino acid positions 362 to 367, 832 to 838 and 858 to 862 in SEQ ID NO: 3) are found in the encoded protein. Subcellular immunodetection of HA-tagged MOM protein confirms its nuclear localization. Similarity to the actin binding domain of chicken tensin (amino acid position 1899 to 1941 in SEQ ID NO: 3) and a predicted membrane spanning domain (amino acid position 995 to 1015 in SEQ ID NO: 3) are also detected. Additionally, the encoded protein contains three types of repetitive regions or internal repeats essentially defined by amino acid positions 177 to 350, 1462 to 1672 and 1848 to 1894 OF SEQ ID NO: 3.

Example 6: ***Homologous genes in other species***

A putative proline/hydroxyproline-rich glycoprotein of *Arabidopsis thaliana* showing partial similarity to the MOM protein is disclosed as GenBank accession nubmer AAD29829). The similarity is 34-47% depending on the region and is only seen in the second half of the MOM protein (i.e. amino acids 1368 to 1944).

The MOM cDNA clone is used to probe genomic DNA from turnip, tomato, tobacco, maize, mouse, fruit fly and man for the presence of homologous genes by Southern blot analysis. Hybridization under conditions of low stringency is found in all cases. Cross-hybridizing clones from libraries can be identified and sequenced.

A genomic library of the *Brassica oleracea* var. acephala (obtained from Dr. Mark Cock, INRA, CNRS, Lyon, France) are screened with the *MOM* cDNA under stringent conditions. Two positive clones are obtained, subcloned, and partially sequenced. Partial sequences of clone 1 show similarity to different regions in the *MOM* gene (80-86% at DNA level and 62-80% at amino acid level) which encode the N-terminal, ATPase, and C-terminal parts of the MOM protein. All three putative nuclear localization sequences of the MOM protein are fully conserved in clone 1. Partial sequences of clone 2 also show similarity regions in the *MOM* gene (64-76% at DNA level and 55-64% at amino acid level) which encode the ATPase, putative transmembrane, and C-terminal parts of the MOM protein. The sequences of clones 1 and 2 are not identical, suggesting the presence of, at least, two homologous genes in *Brassica oleracea.* Examples of partial sequences obtained from clone 1 and 2 are given in SEQ ID Nos: 26-33.

Additionally a genomic library of *Brassica rapa* (obtained from Dr. Kinya Toriyama, Tohoku University, Sendai, Japan) is screened with the *MOM* cDNA under stringent conditions. Positive signals hybridizing to both a 5' and a 3' part of the *MOM* cDNA are obtained.

Furthermore, a genomic library of *Petunia hybrida* (obtained from Dr. Jan Kooter, Vrije Universiteit, Amsterdam, The Netherlands) is screened with *MOM* cDNA under less stringent conditions. Positive signals hybridizing to both the 5' and 3' part of the *MOM* cDNA are obtained.

Example 7: ***Manipulating marker gene expression by antisense constructs***

The 2.6 kb cDNA fragment and a 1.8 kb RT-PCR fragment amplified by a nested RT-PCR using primers RT1-1 and RT1-2 for the first PCR and primers RT1-3 and RT1-4 for the second PCR, are each inversely cloned into the multiple cloning site of the binary vector pbarbi53 to generate antisense RNA. pbarbi53 is a modified vector of p1'barbi and carries

an expression cassette consisting of the 35S promoter of cauliflower mosaic virus, a multiple cloning site containing Xho I, SnaBI, Hpa I and Cla I restriction sites and the 35S terminator of cauliflower mosaic virus at the HindIII site of p1'barbi. The resulting recombinant plasmids are introduced into Agrobacterium as described in Example 1. The transgenic plant line GUS-TS (obtainable from Dr. H. Vaucheret, INRA, Versailles Cedex, France) of Arabidopsis thaliana ecotype Colombia containing a transcriptionally silenced locus with multiple copies of a chimeric beta-glucuronidase (gus) gene, is transformed with the recombinant plasmids as described in Example 1 and transformants are selected as described by Mengiste et al, Plant J 12: 945-948, 1997. pbarbi53 vector DNA is used in control transformations. The transformants are examined for reactivation of the gus gene by histochemical staining. A cotyledon leaf is soaked in gus staining solution (100 mM sodium phosphate buffer (pH 7.0), 0.05% 5-bromo-4-chloro-3-indolyl-beta-D-glucuronidase, 0.1% sodium azide) under vacuum for 10 min and then incubated at 37ºC overnight. While strong gus activity is observed in the plants transformed with the recombinant plasmid carrying the 2.6 kb cDNA, plants transformed with the recombinant plasmid carrying the 1.8 kb RT-PCR fragment or pbarbi53 do not show any gus activity above background. Therefore, expression of the antisense RNA of the 2.6 kb cDNA mimicks the mutant phenotype and confirms that sequences shown in SEQ ID NO: 1, SEQ ID NO: 2 and SEQ ID NO: 3 represent the genetic information for a component of the transcriptional gene silencing system.

What is claimed is:

1. DNA comprising an open reading frame encoding a protein characterized by an amino acid sequence comprising a component sequence of at least 150 amino acid residues having 40% or more identity with an aligned component sequence of SEQ ID NO: 3.

2. The DNA according to claim 1 comprising an open reading frame encoding a protein having the formula R_1-R_2-R_3, wherein
   - R_1, R_2 and R_3 constitute component sequences consisting of amino acid residues independently selected from the group of the amino acid residues Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg, and His,
   - R_1 and R_3 consist independently of 0 to 3000 amino acid residues;
   - R_2 consists of at least 150 amino acid residues; and
   - R_2 is at least 40% identical to an aligned component sequence of SEQ ID NO: 3.

3. The DNA according to claim 1 comprising an open reading frame encoding one or more SWI2/SNF2-like ATPase/helicase motifs.

4. The DNA according to claim 1 comprising an open reading frame encoding a protein having a component sequence defined by amino acids 478-490, 584-600, 617-630, 654-668, 676-690, 718-734, 776-788, 1222-1233, 1738-1749 or 1761-1770 of SEQ ID NO: 3.

5. The DNA according to claim 1, wherein the open reading frame encodes a protein characterized by the amino acid sequence of SEQ ID NO: 3, an allelic amino acid sequence having amino acid residue K instead of M at position 705 of SEQ ID NO: 3, or an amino acid residue D instead of E at position 1219 of SEQ ID NO: 3.

6. The DNA according to claim 1 characterized by the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 2.

7. The DNA according to claim 1, ~~wherein expression of RNA, complementary to mRNA transcribed therefrom,~~ characterized in that expression of corresponding anti-sense RNA in a cell releases silencing of a transgenic marker gene.

8. The protein encoded by the open reading frame of any one of claims 1 to 7.

AMENDED SHEET

9. A method of producing DNA according to claim 1, comprising
   - screening a DNA library for clones which are capable of hybridizing to a fragment of the DNA defined by SEQ ID NO: 1 or SEQ ID NO: 2, wherein said fragment has a length of at least 15 nucleotides;
   - sequencing hybridizing clones;
   - purifying vector DNA of clones comprising an open reading frame encoding a protein characterized by an amino acid sequence comprising a component sequence of at least 150 amino acid residues having 40% or more sequence identity to SEQ ID NO: 3
   - optionally further processing the purified DNA.

10. A polymerase chain reaction wherein at least one oligonucleotide used comprises a sequence of nucleotides which represents 15 or more basepairs of SEQ ID NO: 1 or SEQ ID NO: 2.

10018929 .122101 ...59US

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number
WO 01/00801 A2

(51) International Patent Classification[7]: C12N 15/00

(21) International Application Number: PCT/EP00/05761

(22) International Filing Date: 21 June 2000 (21.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9914623.5 23 June 1999 (23.06.1999) GB

(71) Applicant *(for all designated States except AT, US)*: **NOVARTIS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH).

(71) Applicant *(for AT only)*: **NOVARTIS-ERFINDUNGEN VERWALTUNGSGESELLSCHAFT M.B.H.** [AT/AT]; Brunner Strasse 59, A-1230 Vienna (AT).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **HABU, Yoshiki** [JP/CH]; Im Niederholzboden 54, CH-4125 Riehen (CH). **MITTELSTEN SCHEID, Ortrun** [DE/CH]; Martinskirchplatz 15, CH-4051 Basel (CH). **AMEDEO, Paolo** [IT/CH]; Eimeldingerweg 45, CH-4057 Basel (CH). **PASZKOWSKI, Jerzy** [PL/CH]; Blauenweg 10, CH-4224 Nenzlingen (CH).

(74) Agent: **BECKER, Konrad**; Novartis AG, Patent and Trademark Dept., Agribusiness, Site Rosental, CH-4002 Basel (CH).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/00801 A2

(54) Title: GENE INVOLVED IN EPIGENETIC GENE SILENCING

(57) Abstract: The present invention relates to DNA which encodes proteins involved in gene silencing. Related genes encoding proteins characterized by an amino acid sequence comprising a component sequence of at least 150 amino acid residues having 40% or more identity with an alinged component sequence of SEQ ID NO:3 can be isolated from different sources such as mammalian or plant cells. Further disclosed is a method for isolating DNA according to the invention.

US Case **S -31005A**

DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

☒ Original ☐ Supplemental ☐ Substitute

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

GENE INVOLVED IN EPIGENETIC GENE SILENCING

the specification of which:

☐ is attached hereto.

☐ was filed on ________________ as Application No. ________________
(day/month/year)

and, if this box (☐) contains an ×

☐ was amended on ________________
(day/month/year)

☒ was filed as Patent Cooperation Treaty international Application No.

PCT/EP00/05761 on **21.06.00**
(day/month/year)

and, if this box (☐) contains an ×

☐ entered the national stage in the United States and was accorded Application No.

____

and, if this box (☐) contains an ×

☐ was amended, subsequent to entry into the national stage, on ________________
(day/month/year)

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment(s) referred to above and, if this application was filed as a Patent Cooperation Treaty international application, by any amendments made during the international stage (including any made under Patent Cooperation Treaty Rule 91, Article 19 and Article 34).

I acknowledge my duty to disclose all information which is known by me to be material to the patentability of this application as defined in 37 C.F.R. § 1.56.

US Case **S -31005A**

I hereby claim the benefit under 35 U.S.C. §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate listed below and under 35 U.S.C. §365(a) of any Patent Cooperation Treaty international application(s) designating at least one country other than the United States listed below and have also listed below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application the priority of which is claimed for that subject matter:

COUNTRY/REGION (OR P.C.T.)	APPLICATION No.	FILING DATE (day/month/year)	PRIORITY CLAIMED
GB	**9914623.5**	**23.06.99**	☒ Yes ☐ No
			☐ Yes ☐ No
			☐ Yes ☐ No
			☐ Yes ☐ No
			☐ Yes ☐ No

I hereby claim the benefit under 35 U.S.C. § 119 (e) of any United States provisional application(s) listed below:

APPLICATION NO.	FILING DATE (day/month/year)

I hereby claim the benefit under 35 U.S.C. §120 of any United States application(s) listed below and under 35 U.S.C. §365(c) of any Patent Cooperation Treaty international application(s) designating the United States listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of 35 U.S.C. §112, I acknowledge my duty to disclose all information known to me to be material to patentability as defined in 37 C.F.R. §1.56 which became available between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

United States Application No.	United States Filing Date (day/month/year)	Status (Pending, Abandoned or U.S. Patent No.)	International Application No.	and Filing Date (day/month/year)

US Case **S -31005A**

I hereby appoint the registered practitioners associated with Customer No. 022847, respectively and individually, as my attorneys and agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

If these brackets contain an X [X], I hereby authorize the registered practitioners associated with Customer No. 022847 and any others acting on my behalf to take any action relating to this application based on communications from the Patents and Trademarks Division of Syngenta Crop Protection AG, Basle, Switzerland, or an affiliate thereof or a successor thereto, without direct communication from me.

Please address all communications to the address associated with Customer No. 022847, which is currently Larry W. Stults, Syngenta, Patent and Trademark Department, 3054 Cornwallis Road, Research Triangle Park, NC 27709.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001 and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Full name of sole or first joint inventor	**Yoshiki HABU**		
Inventor's signature	Yoshiki Habu	Date	20.11.2001 (day/month/year)
Residence	**4125 Riehen, Switzerland** CHX		
Citizenship	**Japanese**		
Post Office Address	**Gstaltenrainweg 71** **4125 Riehen** **CH**		

IMPORTANT: Before this declaration is signed, the patent application (the specification, the claims and this declaration) must be read and understood by each person signing it, and no changes may be made in the application after this declaration has been signed.

10038929 .122101

US Case S -31005A

Full name of second joint inventor, if any	Ortrun MITTELSTEN SCHEID
Inventor's signature	[signature] Date 20.11.2001 (day/month/year)
Residence	4051 Basel, Switzerland CHX
Citizenship	German
Post Office Address	Martinskirchplatz 15 4051 Basel CH

2-00

Full name of third joint inventor, if any	Paolo AMEDEO
Inventor's signature	Date (day/month/year)
Residence	San Diego, CA 92121-1125, USA CA
Citizenship	Italian
Post Office Address	Torrey Mesa Research Institute Inc. 3115 Merryfield Row San Diego, CA 92121-1125 USA

3-00

Full name of fourth joint inventor, if any	Jerzy PASZKOWSKI
Inventor's signature	[signature] Date 20.11.2001 (day/month/year)
Residence	4224 Nenzlingen, Switzerland CHX
Citizenship	Polish
Post Office Address	Blauenweg 10 4224 Nenzlingen CH

4-00

US Case **S -31005A**

DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

☒ Original ☐ Supplemental ☐ Substitute

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

GENE INVOLVED IN EPIGENETIC GENE SILENCING

the specification of which:

☐ is attached hereto.

☐ was filed on ______________ as Application No. ______________
(day/month/year)

and, if this box (☐) contains an ×

☐ was amended on ______________
(day/month/year)

☒ was filed as Patent Cooperation Treaty international Application No.

PCT/EP00/05761 on **21.06.00**
(day/month/year)

and, if this box (☐) contains an ×

☐ entered the national stage in the United States and was accorded Application No.

____

and, if this box (☐) contains an ×

☐ was amended, subsequent to entry into the national stage, on ______________
(day/month/year)

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment(s) referred to above and, if this application was filed as a Patent Cooperation Treaty international application, by any amendments made during the international stage (including any made under Patent Cooperation Treaty Rule 91, Article 19 and Article 34).

I acknowledge my duty to disclose all information which is known by me to be material to the patentability of this application as defined in 37 C.F.R. § 1.56.

US Case **S -31005A**

I hereby claim the benefit under 35 U.S.C. §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate listed below and under 35 U.S.C. §365(a) of any Patent Cooperation Treaty international application(s) designating at least one country other than the United States listed below and have also listed below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application the priority of which is claimed for that subject matter:

COUNTRY/REGION (OR P.C.T.)	APPLICATION No.	FILING DATE (day/month/year)	PRIORITY CLAIMED	
GB	**9914623.5**	**23.06.99**	☒ Yes	☐ No
			☐ Yes	☐ No
			☐ Yes	☐ No
			☐ Yes	☐ No
			☐ Yes	☐ No

I hereby claim the benefit under 35 U.S.C. § 119 (e) of any United States provisional application(s) listed below:

APPLICATION NO.	FILING DATE (day/month/year)

I hereby claim the benefit under 35 U.S.C. §120 of any United States application(s) listed below and under 35 U.S.C. §365(c) of any Patent Cooperation Treaty international application(s) designating the United States listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of 35 U.S.C. §112, I acknowledge my duty to disclose all information known to me to be material to patentability as defined in 37 C.F.R. §1.56 which became available between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

United States Application No.	United States Filing Date (day/month/year)	Status (Pending, Abandoned or U.S. Patent No.)	International Application No.	and Filing Date (day/month/year)

US Case **S -31005A**

I hereby appoint the registered practitioners associated with Customer No. 022847, respectively and individually, as my attorneys and agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

If these brackets contain an X [X], I hereby authorize the registered practitioners associated with Customer No. 022847 and any others acting on my behalf to take any action relating to this application based on communications from the Patents and Trademarks Division of Syngenta Crop Protection AG, Basle, Switzerland, or an affiliate thereof or a successor thereto, without direct communication from me.

Please address all communications to the address associated with Customer No. 022847, which is currently Larry W. Stults, Syngenta, Patent and Trademark Department, 3054 Cornwallis Road, Research Triangle Park, NC 27709.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001 and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Full name of sole or first joint inventor **Yoshiki HABU**

Inventor's signature ____________________ Date ____________ (day/month/year)

Residence **4125 Riehen, Switzerland**

Citizenship **Japanese**

Post Office Address **Gstaltenrainweg 71**
4125 Riehen
CH

IMPORTANT: Before this declaration is signed, the patent application (the specification, the claims and this declaration) must be read and understood by each person signing it, and no changes may be made in the application after this declaration has been signed.

US Case S -31005A

Full name of second joint inventor, if any	**Ortrun MITTELSTEN SCHEID**
Inventor's signature	Date (day/month/year)
Residence	**4051 Basel, Switzerland**
Citizenship	**German**
Post Office Address	**Martinskirchplatz 15** **4051 Basel** **CH**

Full name of third joint inventor, if any	**Paolo AMEDEO**
Inventor's signature	Date 05/12/2001 (day/month/year)
Residence	**San Diego, CA 92121-1125, USA**
Citizenship	**Italian**
Post Office Address	**Torrey Mesa Research Institute Inc.** **3115 Merryfield Row** **San Diego, CA 92121-1125** **USA**

Full name of fourth joint inventor, if any	**Jerzy PASZKOWSKI**
Inventor's signature	Date (day/month/year)
Residence	**4224 Nenzlingen, Switzerland**
Citizenship	**Polish**
Post Office Address	**Blauenweg 10** **4224 Nenzlingen** **CH**

SEQUENCE LISTING

<110> Novartis AG
Novartis Research Foundation

<120> Gene involved in epigenetic gene silencing

<130> S-31005A

<140>
<141>

<150> GB 9914623.5
<151> 1999-06-23

<160> 33

<170> PatentIn Ver. 2.1

<210> 1
<211> 10329
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> intron
<222> (1009)..(1295)

<220>
<221> intron
<222> (2551)..(2673)

<220>
<221> intron
<222> (2753)..(2867)

<220>
<221> intron
<222> (3114)..(3506)

<220>
<221> intron
<222> (3681)..(3973)

<220>
<221> intron
<222> (4896)..(4975)

<220>
<221> intron
<222> (5218)..(5777)

<220>
<221> intron

<222> (5883)..(6082)

<220>
<221> intron
<222> (7481)..(7615)

<220>
<221> intron
<222> (7772)..(7914)

<220>
<221> intron
<222> (8071)..(8153)

<220>
<221> intron
<222> (8319)..(8451)

<220>
<221> intron
<222> (8630)..(8718)

<220>
<221> intron
<222> (8919)..(9000)

<220>
<221> intron
<222> (9212)..(9284)

<400> 1
aatatttaag tttggtttat attctttcta gtaatctttg aaatattgta agagataatg 60
cttctaataa ataacattgg atttattgga attaatgtat tgaaaaaact atgcaaatac 120
tacagtgtat tttggaacga ccaaaatgat atatgtaaac tttcgttcta gtcttctaca 180
tagtgtaata ggatagcgga caaggttgat cgactctaaa cattatgggt acgtaattcc 240
gcagtggtta cagtctactg tcgaggccaa actggtaatt aaacgtttga agtttagaga 300
aatattttga tgatgagtac cacaatcaaa gatgataggt gttaatcact gtaaaaatgt 360
tgattgaata ctacgaatgc agaacatata catattttta atctctttgg aatttttgtt 420
tttgttttta tcatttttga atacacgaag agctcagtta tatttcatat tgtatatgaa 480
tttgttctat ttaatcttca attctagcaa catactctta tgctaattcg tttcatattt 540
tagtatagta taaaaattac aaatttcaaa acaaactata agtaatatac taacatagtc 600
ggtgtaacat ttcgttaatt tcacataaca tatgttaatt acatatgtac actattttg 660
aagtatttta taacttaaaa tatataaatt taaatctaag aaatcacaag catgagtttt 720
tccttccggt aatcgtaaaa tcaaaaatcg ctcgctcgag aaacgccggt gctagaagag 780
gaaagtaccg tacataatcc tgcgaaccca attctcgtct tcttcaaact cagttttccg 840
aaaccccaaa caccgcgagg attgcatggc ctgaagaacc acttaatcga gaattgtgct 900
ggaattctca aattttccct cgcgttttc tttcacactc tcggaatcgg aaatttccac 960
caagctccgt caagcgatag attctgacaa ttacacactt tcgcgcaggt atgcttcctt 1020
ccctgtttta ggttggtgtt aatctatcgg tgaatcgaag gttttgggcc tcgggctttg 1080
cgttttaggt ttttcagaga atcttatcta cttggggatg gatcttaggc gtttgttaga 1140
tgtaactcat tagttttgca tataggaatt ttgatttgaa agttaggtcg ccggatttgt 1200
agacattttg tttgatggtc ttcttcggtg ctcacattct ttgtttttaa gtgcttgatt 1260
tggttgctaa ggtcctttcc gttgcgtgct ctcagtgaat atgaagaaag atgaaaagat 1320
tggtttgacg gggagaacca tttacaccag atccctagca gcttcaattc ctgcctcagt 1380
tgaacaagaa acccctggtt tgaggaggtc aagccggggg acaccatcta cgaaggtaat 1440

aactccagct tctgctacta gaaagtcaga gagactggct ccctcacctg cttcagtttc 1500
aaaaaagtcc ggtggaatcg tcaagaattc cacaccaagt tctttgcgaa ggtccaatag 1560
ggggaagact gaagtatcct tgcagagttc caaaggatca gataattcta tcaggaaagg 1620
agatacttca ccggatattg agcagagaaa ggatagtgtt gaagagtcga cagataagat 1680
caagcctata atgtcagccc gaagttacag ggcattgttt agagggaagc tcaaggaatc 1740
tgaggcatta gttgatgctt ccccaaatga agaggaacta gtagttgttg gttgttctcg 1800
ccgcatacct gcaggcaatg atgatgttca aggtaaaaca gattgtccac cacctgcaga 1860
tgcaggatca aaaaggctgc cagttgacga aactagtttg gacaagggca ctgattttcc 1920
tttgaaatca gttacggaga ccgagaagat agtgcttgat gcatccccca tagttgaaac 1980
tggggatgac agtgttatag gttcaccatc tgagaattta gagacacaaa agcttcaaga 2040
tggtaagaca gattgttcac cacctgcaaa tgcagaatcg aaaacgctgc cagttggtga 2100
aactagttta gaaaaagaat atccacaaaa gtttcaagat gataacacag attgtctacc 2160
acctgcaaat gcagaatcaa aaaggctgcc agttggcgaa actagtttag aaaaggacac 2220
tgattttcct ttgaaatcaa ctacggagac tggaaagatg gttctttatg catcccccat 2280
agttgaaact agggatgaca gcgttatatg ttcaccatct acaaatttag aaacccaaaa 2340
gcttcttgtc agtaaaactg gcttagaaac cgacatagtt ttgcctttga aaagaaaaag 2400
agacactgca gaaattgagc tggatgcatg tgctacagtt gcaaatggag atgatcacgt 2460
tatgagttct gatggggtca ttccatctcc atctgggtgc aaaaatgata atcgacctga 2520
aatgtgcaac acgtgtaaaa aacggcaaaa gtaagagttt ttttagtgtt gtctgtctat 2580
tgaaacgatc tgccaatgtt gaatgttggg cagatgggtt tgattcttag gatatatgtt 2640
ctgtattgta atgagttgtt caaaattttg aagggtcaac ggtgattgtc aaaataggag 2700
tgtttgctcc tgcattgtcc agccagttga agaatctgat aacgtgacac aggttggttt 2760
ctaattactt tcggagaccc gttaatcagt ggactcttaa atagttagat actagattta 2820
cttatccttt tacttgtaat ctgcaattct attttgcatt tgattaggat atgaaagaaa 2880
ctggaccagt tacgagcaga gaatatgagg agaacgggca aatacaacat ggtaaatcaa 2940
gtgatcccaa attctattct tcggtgtacc cagagtattg ggttcctgtg cagctatcag 3000
atgtacagct ggagcaatac tgtcagactc tcttctccaa atccttatct ctttcttcac 3060
tttcgaagat tgatcttgga gctctagaag aaactctcaa ttctgtaaga aaagtaagtt 3120
acttgatttt aaaaacactt attcttcaat gcacttgtga gttaagtacc cagttattac 3180
tggtgataag ataaagaaag caatagaaaa attgataagg tgttcaccgc attgcagcca 3240
aaaaaacaat tctgtgttcc atgctttcaa gaggttgtca cataggtgtt atgcctttct 3300
gtttgatgtt tggtagagca aaggttttgg gtctatttgt tttatgcttt tttgaaacac 3360
atagaacctg gcaaacttga cagttttggg gttgcttaga tatacgacta ttgtcggtca 3420
gcatcacatt ttctcaaggc ctctttctgc atgttaatgt gtgaatatat taaaatcttc 3480
tttatgtgtt tgcaacttgt tgacagacct gtgaccatcc atacgttatg gatgcatctt 3540
tgaaacaact gctcaccaag aatctggagt tgcatgaaat cctggatgta gaaattaaag 3600
cgagcgggaa acttcacctc cttgataaaa tgcttactca tataaaaaag aatggtttaa 3660
aagcagttgt cttctaccag gtgcattttc tattacttgc gaatgtgaat agctctatgt 3720
ttgtcatgaa tacgtcactt tgtgcattct caatatatgt gcattttctt tttgacaatg 3780
gaattctgtc ttgtattgaa atttgagtgg gatgaaagta tgctttttat cgtgcaatta 3840
tgaagtgtaa gttagccttc agcagtcagc tagcattatg agatatgctg aactaaaatg 3900
tttcttttct cttctttctt tttcgttata tgtgcctcat gtatgtttga attacagttt 3960
ttattttcag caggcaacac aaaccccctga agggcttctg cttggtaata ttctcgaaga 4020
ttttgtgggt caaagatttg gtccaaaatc ttatgagcat gggatatatt cctcaaagaa 4080
gaactccgct ataaacaatt tcaacaagga gagtcaatgc tgtgttctgc tgttggaaac 4140
acgtgcctgc agtcaaacca ttaaactctt gcgagctgat gcgtttattc tttttggaag 4200
cagcttgaat ccatcgcatg atgttaagca cgtagagaag ataaaaatcg agtcatgttc 4260
tgaaagaact aagatattcc gattgtactc agtatgtaca gttgaagaaa aagccctgat 4320
tctggctagg caaaatatgc ggcaaaataa agctgtagag aacctaaacc gctctctcac 4380
gcacgcactg ctcatgtggg gggcgtcata cttatttgat aaactggatc attttcacag 4440
cagtgaaact ccagattcag gagtttcatt tgaacaatct attatggacg gcgtgattca 4500
tgaattctcg tccatacttt cttccaaagg tggagaagaa aatgaagtca agctgtgtct 4560
acttttggag gccaagcatg ctcagggaac ttacagcagt gattctactc tatttggtga 4620
agaccatatt aagttgtcag atgaagagag tccaaatata ttttggtcaa agctgttggg 4680
gggaaaaaat cctatgtgga aatacccttc agatactccc caaaggaatc gaaaacgagt 4740

tcagtatttt gagggttctg aagcgagtcc caaaactggc gatggtggaa atgcaaagaa 4800
gcgaaagaag gcttctgatg atgtcactga tccccgggtc actgatccgc cagtagatga 4860
tgatgaaaga aaggcctctg ggaaggatca catgggtaaa atagtttaat ttctgctccg 4920
atacctctag tgttcattga ttatgcaact actttgctga ctatctttcc tacaggggct 4980
ttggagtcac caaaagtcat aacactccag tcatcatgta aatcttctgg tacagatggt 5040
acattggatg gaaatgatgc ttttggcttg tattctatgg gcagccatat ctctggaatc 5100
ccagaggata tgttagctag tcaagattgg gggaaaatac cggatgaatc acagaggagg 5160
ctccacactg ttttaaagcc gaagatggca aaactttgcc aagttttgca tctttcagta 5220
agtggccttt ttcacctcca caacttattt tagccttgca tatgcttata tatagctgat 5280
tgcaactgta gttgttacct gatttcctgt tacagccaaa tgtgagagtt ttattcttca 5340
actatatcca tccgtttaag catattttat ttcttatatc tggcttcgtt accaatgcac 5400
tgttaaaatg agcaactgct gcacaaaaca gtaggtagtt atgtgcctca tgtcattcat 5460
tgtttattga agcaaagaaa tttctgtcta ctttacatga tccatctgtg ggagtatata 5520
actatatata accttaggcc tttgtacctg gctgatcaaa gacatgtcaa aagtttatct 5580
gttcgctgtt ggtatagaaa ctaatacagt gtctgatgct attttaaggt agtcttatgt 5640
cttcacatat tggctaatag atgtttccgc tgtcgtgtcc atatacttct gtgattatca 5700
cggtgctccg tctatcaaaa ttgtactaaa aggtattttg caatgtgtga ttggttaaca 5760
gattattttg ttttcaggat gcttgcacaa gcatggtcgg aaattttctc gaatatgtta 5820
ttgaaaatca ccgaatctac gaagagccag ccactacttt tcaggcattc cagatagccc 5880
tggtatgaca gcatttactt tgataattta tgcattgttt ccttcatcat ctgcctttgt 5940
ttagaatgtc ctcagaaggc agcactcctt tagttttaac tttccaatca taggattcaa 6000
atatccatta actggccttt gatcgctgca taatatatga atagttgaca tactgaatac 6060
gttgttaata atgcattttc agagttggat tgcagccttg ttggtaaagc aaattcttag 6120
ccacaaagaa tctctggtcc gtgcaaattc tgaattagct ttcaaatgct ctagagtaga 6180
ggtggattat atttattcga tattgtcctg catgaagagt ctgttcctgg agcatacaca 6240
aggtttgcag ttcgattgct ttggtactaa ttctaaacag tcagtggtta gcacaaaact 6300
agtaaatgaa agtctctcag ggctacagt gcgtgacgaa aagattaata cgaagtcgat 6360
gcgaaatagc tcagaggatg aagagtgcat gactgagaag agatgtagcc attatagcac 6420
agcaacaaga gatatcgaaa agactattag tggcataaaa aagaaataca agaagcaagt 6480
gcaaaagctt gtacaagagc atgaggaaaa gaaaatggag ctgttaaata tgtatgcaga 6540
caagaagcag aaacttgaaa ctagtaaaag tgtggaagca gcagtaattc gtattacctg 6600
ttcacggacc agtactcaag tgggtgatct caaactgctg gatcataatt atgaaagaaa 6660
gtttgatgaa atcaaaagtg agaaaaatga atgcctcaaa agtctggagc aaatgcacga 6720
ggttgcaaag aagaagttgg ctgaggatga agcctgttgg attaatcgga taaagagctg 6780
ggcagctaaa ttaaaagttt gtgttcccat tcaaagtggc aataacaagc attttagtgg 6840
ttcatcaaac atttcccaaa atgctcctga tgtacaaatt tgcaataatg ctaacgttga 6900
agctacttac gctgatacga attgcatggc ttccaaggtt aatcaagtgc cagaagcaga 6960
aaacacatta ggaaccatgt cgggtggcag cactcaacaa gttcatgaaa tggtggatgt 7020
aagaaatgac gagacaatgg atgtctcagc tttgtctcgt gaacagctta caaagagcca 7080
gtccaatgag cacgcttcta tcactgtgcc tgagattttg attcctgctg actgtcaaga 7140
ggaatttgcg gccttgaacg tgcatttgtc agaagaccag aattgtgaca gaataacatc 7200
tgcggcatca gatgaagatg tttcatcaag ggtgccagag gtatcccagt cactcgaaaa 7260
tctttctgcc tcccccgagt tttctctaaa tagagaggag gctttggtta caacagaaaa 7320
tagaagaaca agtcatgtgg gttttgatac tgataacatt ttggaccagc agaatagaga 7380
agattgttct cttgaccaag agattcctga cgagttagcg atgcctgtgc aacatcttgc 7440
gtctgtggta gagactaggg gtgctgctga atctgatcag gtacttactg gccctgtaga 7500
atagttgatg ccttgttcat ttaatctttt ctaatgttca ttcttgcttt cttgaaaata 7560
acgggtagtg atcagatgtc tttttttctc ttattaaatt cacttttctg gacagtatgg 7620
tcaagatata tgtcctatgc cttcttcact ggctggaaag caacctgacc cagcagcaaa 7680
cactgagagc gaaaatcttg aagaagcaat tgagcctcag tctgctggtt cagaaacagt 7740
agagactact gattttgctg catcacatca ggtccctatt gaagactttc cttttttact 7800
agtttaaagt tatcaatctg tgttatgttc attctaagtt tccgtgagaa aaaggtgggg 7860
aaatgtggtt actgatcaag tctcgttgtt gttttaaatc gactcttttg acagggtgat 7920
caagttacat gtcctttgct atcttcaccg actggaaatc agcctgcgcc agaagcaaat 7980
attgaaggcc aaaatatcaa cacatcagct gagccccatg tagcgggtcc agatgcagta 8040

gagagtggtg attatgcagt aatagatcag gttattgcct taactaaaga caaatgtctt 8100
ttgttgttta aaagtcttac atctttgtaa tgctcgttct ggatatcctg caggaaacaa 8160
tgggtgctca ggatgcatgc tctctgccat ctggatcggt tggaactcag tctgacctag 8220
gagcaaacat tgagggtcaa aatgtcacaa cagtggctca acttcccaca gatggatcag 8280
atgcagttgt aaccggtgga tctcctgtat cagatcaggt acctgcctct gctcaaggac 8340
tttcttatgt gttggtttaa aggtctagtc cttagtaatg ttgaaactaa gcaaacagtg 8400
gatagtgatc atatggttat ttttgcttgt gaatttaata tttctggaca gtgtgcccag 8460
gatgcatctc ctatgccatt atcttcgcct ggaaatcacc ctgatacagc agttaatatc 8520
gagggtttag ataacacatc agtagctgag cctcatataa gtggatcaga tgcatgtgaa 8580
atggaaattt cagaacctgg tccccaagta gagcggtcaa cctttgcaag tcagtaactg 8640
ccttgggcat ttttaagtat cacctaggtc gacatatgtg attgccaaac agctaacaag 8700
gagatgcctt ttgtgcagat cttttccatg aaggtggcgt ggagcattca gcaggtgtaa 8760
cagctcttgt tccatcactt cttaacaatg gtacggaaca gattgccgtt caacctgttc 8820
ctcaaatacc tttccctgtg ttcaacgacc cgtttctgca tgaactggag aagttgcgga 8880
gagaatcaga gaactcaaag aagacttttg aagaaaaagt cagtttccct cattacccag 8940
ttacctcttg ttttggttta ttttctagct gcccattgac tctcagttgc ttgtgagcag 9000
aaatcaatct tgaaagctga actcgagagg aagatggctg aagtacaagc agagtttcga 9060
agaaaatttc atgaggtaga agccgagcat aacaccagaa cgacaaagat agagaaggat 9120
aagaatcttg ttataatgaa caaactgttg gcgaatgcgt tcttgtccaa atgtactgac 9180
aagaaggtat ctccctcagg agctccaagg ggtaagtgtc gaataatata gcaaattggt 9240
tttaaaaata aggcgacgaa gtcataatag cactttttct ccaggtaaaa ttcagcagct 9300
agcacagaga gcagcacaag tgagtgcact gagaaattac attgctcctc agcagcttca 9360
ggcatcttct tttcctgctc ctgctctggt ttcggctcct ctgcaacttc agcaatcatc 9420
atttcctgct cctggtccgg ctcctctgca gcctcaggca tcttcgtttc cttcttcagt 9480
ctctcgtcca tcagcccttc ttctgaattt tgcggtctgt ccaatgcctc agcccagaca 9540
gcctctcata tccaacatag ctccaactcc atcagttact cctgcaacaa atccaggtct 9600
gcgttctcct gcaccacacc taaactcata tagaccatcc tcttcaactc ccgtcgccac 9660
agctactcca acctcgtcag tgcctcctca agctttgaca tattcagctg tgtcaattca 9720
gcagcagcaa gaacaacaac cgcaacagag cttgagcagt ggattgcaga gcaacaatga 9780
agtggtttgt ctttctgacg acgagtgacc taagaggaga gatggttagg gtcttagtta 9840
ttgattttta gagagttaat aatagtatat atatatatgt ataagtaggt tacctaatct 9900
ctgtcgttaa tctaatttag tgagtcagga accgactcgt tggctaaggt ctctcctttt 9960
gaaacgcaac gttctacttt catgtatata aatacagtct gatcacacaa cacaaattga 10020
tgattgaaaa tactactgat ttaactttat agaaaaccca aattatagag cgacaacttt 10080
ataaacatgt caaacttcga agttaaaatt taagacccca taattttaca attatagatt 10140
ttaatactcc aactattttg tgatgttaaa agaagtatcc gagtcttttc tttccagttt 10200
ccccaccgtc ccatgactcc cccagccagt agaaaaagcc aaaaaagtaa acaaaaagtc 10260
gttaaaaaag ttaaattaaa aaaaaaatag atagttgacg tttactaaag tgatttgaat 10320
tgaacaatt 10329

<210> 2
<211> 6571
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (310)..(6312)

<400> 2
cacaagcatg agtttttcct tccggtaatc gtaaaatcaa aaatcgctcg ctcgagaaac 60

gccggtgcta gaagaggaaa gtaccgtaca taatcctgcg aacccaattc tcgtcttctt 120

```
caaactcagt tttccgaaac cccaaacacc gcgaggattg catggcctga agaaccactt 180

aatcgagaat tgtgctggaa ttctcaaatt ttccctcgcg tttttctttc acactctcgg 240

aatcggaaat ttccaccaag ctccgtcaag cgatagattc tgacaattac acactttcgc 300

gcagtgaat atg aag aaa gat gaa aag att ggt ttg acg ggg aga acc att 351
          Met Lys Lys Asp Glu Lys Ile Gly Leu Thr Gly Arg Thr Ile
           1               5                  10

tac acc aga tcc cta gca gct tca att cct gcc tca gtt gaa caa gaa    399
Tyr Thr Arg Ser Leu Ala Ala Ser Ile Pro Ala Ser Val Glu Gln Glu
 15                  20                  25                  30

acc cct ggt ttg agg agg tca agc cgg ggg aca cca tct acg aag gta    447
Thr Pro Gly Leu Arg Arg Ser Ser Arg Gly Thr Pro Ser Thr Lys Val
                 35                  40                  45

ata act cca gct tct gct act aga aag tca gag aga ctg gct ccc tca    495
Ile Thr Pro Ala Ser Ala Thr Arg Lys Ser Glu Arg Leu Ala Pro Ser
             50                  55                  60

cct gct tca gtt tca aaa aag tcc ggt gga atc gtc aag aat tcc aca    543
Pro Ala Ser Val Ser Lys Lys Ser Gly Gly Ile Val Lys Asn Ser Thr
         65                  70                  75

cca agt tct ttg cga agg tcc aat agg ggg aag act gaa gta tcc ttg    591
Pro Ser Ser Leu Arg Arg Ser Asn Arg Gly Lys Thr Glu Val Ser Leu
     80                  85                  90

cag agt tcc aaa gga tca gat aat tct atc agg aaa gga gat act tca    639
Gln Ser Ser Lys Gly Ser Asp Asn Ser Ile Arg Lys Gly Asp Thr Ser
 95                 100                 105                 110

ccg gat att gag cag aga aag gat agt gtt gaa gag tcg aca gat aag    687
Pro Asp Ile Glu Gln Arg Lys Asp Ser Val Glu Glu Ser Thr Asp Lys
                115                 120                 125

atc aag cct ata atg tca gcc cga agt tac agg gca ttg ttt aga ggg    735
Ile Lys Pro Ile Met Ser Ala Arg Ser Tyr Arg Ala Leu Phe Arg Gly
            130                 135                 140

aag ctc aag gaa tct gag gca tta gtt gat gct tcc cca aat gaa gag    783
Lys Leu Lys Glu Ser Glu Ala Leu Val Asp Ala Ser Pro Asn Glu Glu
        145                 150                 155

gaa cta gta gtt gtt ggt tgt tct cgc cgc ata cct gca ggc aat gat    831
Glu Leu Val Val Val Gly Cys Ser Arg Arg Ile Pro Ala Gly Asn Asp
    160                 165                 170

gat gtt caa ggt aaa aca gat tgt cca cca cct gca gat gca gga tca    879
Asp Val Gln Gly Lys Thr Asp Cys Pro Pro Pro Ala Asp Ala Gly Ser
175                 180                 185                 190

aaa agg ctg cca gtt gac gaa act agt ttg gac aag ggc act gat ttt    927
```

```
Lys Arg Leu Pro Val Asp Glu Thr Ser Leu Asp Lys Gly Thr Asp Phe
                195                 200                 205

cct ttg aaa tca gtt acg gag acc gag aag ata gtg ctt gat gca tcc      975
Pro Leu Lys Ser Val Thr Glu Thr Glu Lys Ile Val Leu Asp Ala Ser
            210                 215                 220

ccc ata gtt gaa act ggg gat gac agt gtt ata ggt tca cca tct gag     1023
Pro Ile Val Glu Thr Gly Asp Asp Ser Val Ile Gly Ser Pro Ser Glu
        225                 230                 235

aat tta gag aca caa aag ctt caa gat ggt aag aca gat tgt tca cca     1071
Asn Leu Glu Thr Gln Lys Leu Gln Asp Gly Lys Thr Asp Cys Ser Pro
    240                 245                 250

cct gca aat gca gaa tcg aaa acg ctg cca gtt ggt gaa act agt tta     1119
Pro Ala Asn Ala Glu Ser Lys Thr Leu Pro Val Gly Glu Thr Ser Leu
255                 260                 265                 270

gaa aaa gaa tat cca caa aag ttt caa gat gat aac aca gat tgt cta     1167
Glu Lys Glu Tyr Pro Gln Lys Phe Gln Asp Asp Asn Thr Asp Cys Leu
                275                 280                 285

cca cct gca aat gca gaa tca aaa agg ctg cca gtt ggc gaa act agt     1215
Pro Pro Ala Asn Ala Glu Ser Lys Arg Leu Pro Val Gly Glu Thr Ser
            290                 295                 300

tta gaa aag gac act gat ttt cct ttg aaa tca act acg gag act gga     1263
Leu Glu Lys Asp Thr Asp Phe Pro Leu Lys Ser Thr Thr Glu Thr Gly
        305                 310                 315

aag atg gtt ctt tat gca tcc ccc ata gtt gaa act agg gat gac agc     1311
Lys Met Val Leu Tyr Ala Ser Pro Ile Val Glu Thr Arg Asp Asp Ser
    320                 325                 330

gtt ata tgt tca cca tct aca aat tta gaa acc caa aag ctt ctt gtc     1359
Val Ile Cys Ser Pro Ser Thr Asn Leu Glu Thr Gln Lys Leu Leu Val
335                 340                 345                 350

agt aaa act ggc tta gaa acc gac ata gtt ttg cct ttg aaa aga aaa     1407
Ser Lys Thr Gly Leu Glu Thr Asp Ile Val Leu Pro Leu Lys Arg Lys
                355                 360                 365

aga gac act gca gaa att gag ctg gat gca tgt gct aca gtt gca aat     1455
Arg Asp Thr Ala Glu Ile Glu Leu Asp Ala Cys Ala Thr Val Ala Asn
            370                 375                 380

gga gat gat cac gtt atg agt tct gat ggg gtc att cca tct cca tct     1503
Gly Asp Asp His Val Met Ser Ser Asp Gly Val Ile Pro Ser Pro Ser
        385                 390                 395

ggg tgc aaa aat gat aat cga cct gaa atg tgc aac acg tgt aaa aaa     1551
Gly Cys Lys Asn Asp Asn Arg Pro Glu Met Cys Asn Thr Cys Lys Lys
    400                 405                 410
```

```
cgg caa aag gtc aac ggt gat tgt caa aat agg agt gtt tgc tcc tgc    1599
Arg Gln Lys Val Asn Gly Asp Cys Gln Asn Arg Ser Val Cys Ser Cys
415                 420                 425                 430

att gtc cag cca gtt gaa gaa tct gat aac gtg aca cag gat atg aaa    1647
Ile Val Gln Pro Val Glu Glu Ser Asp Asn Val Thr Gln Asp Met Lys
                435                 440                 445

gaa act gga cca gtt acg agc aga gaa tat gag gag aac ggg caa ata    1695
Glu Thr Gly Pro Val Thr Ser Arg Glu Tyr Glu Glu Asn Gly Gln Ile
            450                 455                 460

caa cat ggt aaa tca agt gat ccc aaa ttc tat tct tcg gtg tac cca    1743
Gln His Gly Lys Ser Ser Asp Pro Lys Phe Tyr Ser Ser Val Tyr Pro
        465                 470                 475

gag tat tgg gtt cct gtg cag cta tca gat gta cag ctg gag caa tac    1791
Glu Tyr Trp Val Pro Val Gln Leu Ser Asp Val Gln Leu Glu Gln Tyr
    480                 485                 490

tgt cag act ctc ttc tcc aaa tcc tta tct ctt tct tca ctt tcg aag    1839
Cys Gln Thr Leu Phe Ser Lys Ser Leu Ser Leu Ser Ser Leu Ser Lys
495                 500                 505                 510

att gat ctt gga gct cta gaa gaa act ctc aat tct gta aga aaa acc    1887
Ile Asp Leu Gly Ala Leu Glu Glu Thr Leu Asn Ser Val Arg Lys Thr
                515                 520                 525

tgt gac cat cca tac gtt atg gat gca tct ttg aaa caa ctg ctc acc    1935
Cys Asp His Pro Tyr Val Met Asp Ala Ser Leu Lys Gln Leu Leu Thr
            530                 535                 540

aag aat ctg gag ttg cat gaa atc ctg gat gta gaa att aaa gcg agc    1983
Lys Asn Leu Glu Leu His Glu Ile Leu Asp Val Glu Ile Lys Ala Ser
        545                 550                 555

ggg aaa ctt cac ctc ctt gat aaa atg ctt act cat ata aaa aag aat    2031
Gly Lys Leu His Leu Leu Asp Lys Met Leu Thr His Ile Lys Lys Asn
    560                 565                 570

ggt tta aaa gca gtt gtc ttc tac cag gca aca caa acc cct gaa ggg    2079
Gly Leu Lys Ala Val Val Phe Tyr Gln Ala Thr Gln Thr Pro Glu Gly
575                 580                 585                 590

ctt ctg ctt ggt aat att ctc gaa gat ttt gtg ggt caa aga ttt ggt    2127
Leu Leu Leu Gly Asn Ile Leu Glu Asp Phe Val Gly Gln Arg Phe Gly
                595                 600                 605

cca aaa tct tat gag cat ggg ata tat tcc tca aag aag aac tcc gct    2175
Pro Lys Ser Tyr Glu His Gly Ile Tyr Ser Ser Lys Lys Asn Ser Ala
            610                 615                 620

ata aac aat ttc aac aag gag agt caa tgc tgt gtt ctg ctg ttg gaa    2223
Ile Asn Asn Phe Asn Lys Glu Ser Gln Cys Cys Val Leu Leu Leu Glu
        625                 630                 635
```

```
aca cgt gcc tgc agt caa acc att aaa ctc ttg cga gct gat gcg ttt     2271
Thr Arg Ala Cys Ser Gln Thr Ile Lys Leu Leu Arg Ala Asp Ala Phe
    640                 645                 650

att ctt ttt gga agc agc ttg aat cca tcg cat gat gtt aag cac gta     2319
Ile Leu Phe Gly Ser Ser Leu Asn Pro Ser His Asp Val Lys His Val
655                 660                 665                 670

gag aag ata aaa atc gag tca tgt tct gaa aga act aag ata ttc cga     2367
Glu Lys Ile Lys Ile Glu Ser Cys Ser Glu Arg Thr Lys Ile Phe Arg
                675                 680                 685

ttg tac tca gta tgt aca gtt gaa gaa aaa gcc ctg att ctg gct agg     2415
Leu Tyr Ser Val Cys Thr Val Glu Glu Lys Ala Leu Ile Leu Ala Arg
            690                 695                 700

caa aat atg cgg caa aat aaa gct gta gag aac cta aac cgc tct ctc     2463
Gln Asn Met Arg Gln Asn Lys Ala Val Glu Asn Leu Asn Arg Ser Leu
        705                 710                 715

acg cac gca ctg ctc atg tgg ggg gcg tca tac tta ttt gat aaa ctg     2511
Thr His Ala Leu Leu Met Trp Gly Ala Ser Tyr Leu Phe Asp Lys Leu
    720                 725                 730

gat cat ttt cac agc agt gaa act cca gat tca gga gtt tca ttt gaa     2559
Asp His Phe His Ser Ser Glu Thr Pro Asp Ser Gly Val Ser Phe Glu
735                 740                 745                 750

caa tct att atg gac ggc gtg att cat gaa ttc tcg tcc ata ctt tct     2607
Gln Ser Ile Met Asp Gly Val Ile His Glu Phe Ser Ser Ile Leu Ser
                755                 760                 765

tcc aaa ggt gga gaa gaa aat gaa gtc aag ctg tgt cta ctt ttg gag     2655
Ser Lys Gly Gly Glu Glu Asn Glu Val Lys Leu Cys Leu Leu Leu Glu
            770                 775                 780

gcc aag cat gct cag gga act tac agc agt gat tct act cta ttt ggt     2703
Ala Lys His Ala Gln Gly Thr Tyr Ser Ser Asp Ser Thr Leu Phe Gly
        785                 790                 795

gaa gac cat att aag ttg tca gat gaa gag agt cca aat ata ttt tgg     2751
Glu Asp His Ile Lys Leu Ser Asp Glu Glu Ser Pro Asn Ile Phe Trp
    800                 805                 810

tca aag ctg ttg ggg gga aaa aat cct atg tgg aaa tac cct tca gat     2799
Ser Lys Leu Leu Gly Gly Lys Asn Pro Met Trp Lys Tyr Pro Ser Asp
815                 820                 825                 830

act ccc caa agg aat cga aaa cga gtt cag tat ttt gag ggt tct gaa     2847
Thr Pro Gln Arg Asn Arg Lys Arg Val Gln Tyr Phe Glu Gly Ser Glu
                835                 840                 845

gcg agt ccc aaa act ggc gat ggt gga aat gca aag aag cga aag aag     2895
Ala Ser Pro Lys Thr Gly Asp Gly Gly Asn Ala Lys Lys Arg Lys Lys
```

850 855 860

gct tct gat gat gtc act gat ccc cgg gtc act gat ccg cca gta gat 2943
Ala Ser Asp Asp Val Thr Asp Pro Arg Val Thr Asp Pro Pro Val Asp
865 870 875

gat gat gaa aga aag gcc tct ggg aag gat cac atg ggg gct ttg gag 2991
Asp Asp Glu Arg Lys Ala Ser Gly Lys Asp His Met Gly Ala Leu Glu
880 885 890

tca cca aaa gtc ata aca ctc cag tca tca tgt aaa tct tct ggt aca 3039
Ser Pro Lys Val Ile Thr Leu Gln Ser Ser Cys Lys Ser Ser Gly Thr
895 900 905 910

gat ggt aca ttg gat gga aat gat gct ttt ggc ttg tat tct atg ggc 3087
Asp Gly Thr Leu Asp Gly Asn Asp Ala Phe Gly Leu Tyr Ser Met Gly
915 920 925

agc cat atc tct gga atc cca gag gat atg tta gct agt caa gat tgg 3135
Ser His Ile Ser Gly Ile Pro Glu Asp Met Leu Ala Ser Gln Asp Trp
930 935 940

ggg aaa ata ccg gat gaa tca cag agg agg ctc cac act gtt tta aag 3183
Gly Lys Ile Pro Asp Glu Ser Gln Arg Arg Leu His Thr Val Leu Lys
945 950 955

ccg aag atg gca aaa ctt tgc caa gtt ttg cat ctt tca gat gct tgc 3231
Pro Lys Met Ala Lys Leu Cys Gln Val Leu His Leu Ser Asp Ala Cys
960 965 970

aca agc atg gtc gga aat ttt ctc gaa tat gtt att gaa aat cac cga 3279
Thr Ser Met Val Gly Asn Phe Leu Glu Tyr Val Ile Glu Asn His Arg
975 980 985 990

atc tac gaa gag cca gcc act act ttt cag gca ttc cag ata gcc ctg 3327
Ile Tyr Glu Glu Pro Ala Thr Thr Phe Gln Ala Phe Gln Ile Ala Leu
995 1000 1005

agt tgg att gca gcc ttg ttg gta aag caa att ctt agc cac aaa gaa 3375
Ser Trp Ile Ala Ala Leu Leu Val Lys Gln Ile Leu Ser His Lys Glu
1010 1015 1020

tct ctg gtc cgt gca aat tct gaa tta gct ttc aaa tgc tct aga gta 3423
Ser Leu Val Arg Ala Asn Ser Glu Leu Ala Phe Lys Cys Ser Arg Val
1025 1030 1035

gag gtg gat tat att tat tcg ata ttg tcc tgc atg aag agt ctg ttc 3471
Glu Val Asp Tyr Ile Tyr Ser Ile Leu Ser Cys Met Lys Ser Leu Phe
1040 1045 1050

ctg gag cat aca caa ggt ttg cag ttc gat tgc ttt ggt act aat tct 3519
Leu Glu His Thr Gln Gly Leu Gln Phe Asp Cys Phe Gly Thr Asn Ser
1055 1060 1065 1070

aaa cag tca gtg gtt agc aca aaa cta gta aat gaa agt ctc tca ggg 3567

Lys Gln Ser Val Val Ser Thr Lys Leu Val Asn Glu Ser Leu Ser Gly
1075 1080 1085

gct aca gtg cgt gac gaa aag att aat acg aag tcg atg cga aat agc 3615
Ala Thr Val Arg Asp Glu Lys Ile Asn Thr Lys Ser Met Arg Asn Ser
1090 1095 1100

tca gag gat gaa gag tgc atg act gag aag aga tgt agc cat tat agc 3663
Ser Glu Asp Glu Glu Cys Met Thr Glu Lys Arg Cys Ser His Tyr Ser
1105 1110 1115

aca gca aca aga gat atc gaa aag act att agt ggc ata aaa aag aaa 3711
Thr Ala Thr Arg Asp Ile Glu Lys Thr Ile Ser Gly Ile Lys Lys Lys
1120 1125 1130

tac aag aag caa gtg caa aag ctt gta caa gag cat gag gaa aag aaa 3759
Tyr Lys Lys Gln Val Gln Lys Leu Val Gln Glu His Glu Glu Lys Lys
1135 1140 1145 1150

atg gag ctg tta aat atg tat gca gac aag aag cag aaa ctt gaa act 3807
Met Glu Leu Leu Asn Met Tyr Ala Asp Lys Lys Gln Lys Leu Glu Thr
1155 1160 1165

agt aaa agt gtg gaa gca gca gta att cgt att acc tgt tca cgg acc 3855
Ser Lys Ser Val Glu Ala Ala Val Ile Arg Ile Thr Cys Ser Arg Thr
1170 1175 1180

agt act caa gtg ggt gat ctc aaa ctg ctg gat cat aat tat gaa aga 3903
Ser Thr Gln Val Gly Asp Leu Lys Leu Leu Asp His Asn Tyr Glu Arg
1185 1190 1195

aag ttt gat gaa atc aaa agt gag aaa aat gaa tgc ctc aaa agt ctg 3951
Lys Phe Asp Glu Ile Lys Ser Glu Lys Asn Glu Cys Leu Lys Ser Leu
1200 1205 1210

gag caa atg cac gag gtt gca aag aag aag ttg gct gag gat gaa gcc 3999
Glu Gln Met His Glu Val Ala Lys Lys Lys Leu Ala Glu Asp Glu Ala
1215 1220 1225 1230

tgt tgg att aat cgg ata aag agc tgg gca gct aaa tta aaa gtt tgt 4047
Cys Trp Ile Asn Arg Ile Lys Ser Trp Ala Ala Lys Leu Lys Val Cys
1235 1240 1245

gtt ccc att caa agt ggc aat aac aag cat ttt agt ggt tca tca aac 4095
Val Pro Ile Gln Ser Gly Asn Asn Lys His Phe Ser Gly Ser Ser Asn
1250 1255 1260

att tcc caa aat gct cct gat gta caa att tgc aat aat gct aac gtt 4143
Ile Ser Gln Asn Ala Pro Asp Val Gln Ile Cys Asn Asn Ala Asn Val
1265 1270 1275

gaa gct act tac gct gat acg aat tgc atg gct tcc aag gtt aat caa 4191
Glu Ala Thr Tyr Ala Asp Thr Asn Cys Met Ala Ser Lys Val Asn Gln
1280 1285 1290

gtg cca gaa gca gaa aac aca tta gga acc atg tcg ggt ggc agc act 4239
Val Pro Glu Ala Glu Asn Thr Leu Gly Thr Met Ser Gly Gly Ser Thr
1295 1300 1305 1310

caa caa gtt cat gaa atg gtg gat gta aga aat gac gag aca atg gat 4287
Gln Gln Val His Glu Met Val Asp Val Arg Asn Asp Glu Thr Met Asp
1315 1320 1325

gtc tca gct ttg tct cgt gaa cag ctt aca aag agc cag tcc aat gag 4335
Val Ser Ala Leu Ser Arg Glu Gln Leu Thr Lys Ser Gln Ser Asn Glu
1330 1335 1340

cac gct tct atc act gtg cct gag att ttg att cct gct gac tgt caa 4383
His Ala Ser Ile Thr Val Pro Glu Ile Leu Ile Pro Ala Asp Cys Gln
1345 1350 1355

gag gaa ttt gcg gcc ttg aac gtg cat ttg tca gaa gac cag aat tgt 4431
Glu Glu Phe Ala Ala Leu Asn Val His Leu Ser Glu Asp Gln Asn Cys
1360 1365 1370

gac aga ata aca tct gcg gca tca gat gaa gat gtt tca tca agg gtg 4479
Asp Arg Ile Thr Ser Ala Ala Ser Asp Glu Asp Val Ser Ser Arg Val
1375 1380 1385 1390

cca gag gta tcc cag tca ctc gaa aat ctt tct gcc tcc ccc gag ttt 4527
Pro Glu Val Ser Gln Ser Leu Glu Asn Leu Ser Ala Ser Pro Glu Phe
1395 1400 1405

tct cta aat aga gag gag gct ttg gtt aca aca gaa aat aga aga aca 4575
Ser Leu Asn Arg Glu Glu Ala Leu Val Thr Thr Glu Asn Arg Arg Thr
1410 1415 1420

agt cat gtg ggt ttt gat act gat aac att ttg gac cag cag aat aga 4623
Ser His Val Gly Phe Asp Thr Asp Asn Ile Leu Asp Gln Gln Asn Arg
1425 1430 1435

gaa gat tgt tct ctt gac caa gag att cct gac gag tta gcg atg cct 4671
Glu Asp Cys Ser Leu Asp Gln Glu Ile Pro Asp Glu Leu Ala Met Pro
1440 1445 1450

gtg caa cat ctt gcg tct gtg gta gag act agg ggt gct gct gaa tct 4719
Val Gln His Leu Ala Ser Val Val Glu Thr Arg Gly Ala Ala Glu Ser
1455 1460 1465 1470

gat cag tat ggt caa gat ata tgt cct atg cct tct tca ctg gct gga 4767
Asp Gln Tyr Gly Gln Asp Ile Cys Pro Met Pro Ser Ser Leu Ala Gly
1475 1480 1485

aag caa cct gac cca gca gca aac act gag agc gaa aat ctt gaa gaa 4815
Lys Gln Pro Asp Pro Ala Ala Asn Thr Glu Ser Glu Asn Leu Glu Glu
1490 1495 1500

gca att gag cct cag tct gct ggt tca gaa aca gta gag act act gat 4863
Ala Ile Glu Pro Gln Ser Ala Gly Ser Glu Thr Val Glu Thr Thr Asp
1505 1510 1515

ttt gct gca tca cat cag ggt gat caa gtt aca tgt cct ttg cta tct 4911
Phe Ala Ala Ser His Gln Gly Asp Gln Val Thr Cys Pro Leu Leu Ser
1520 1525 1530

tca ccg act gga aat cag cct gcg cca gaa gca aat att gaa ggc caa 4959
Ser Pro Thr Gly Asn Gln Pro Ala Pro Glu Ala Asn Ile Glu Gly Gln
1535 1540 1545 1550

aat atc aac aca tca gct gag ccc cat gta gcg ggt cca gat gca gta 5007
Asn Ile Asn Thr Ser Ala Glu Pro His Val Ala Gly Pro Asp Ala Val
1555 1560 1565

gag agt ggt gat tat gca gta ata gat cag gaa aca atg ggt gct cag 5055
Glu Ser Gly Asp Tyr Ala Val Ile Asp Gln Glu Thr Met Gly Ala Gln
1570 1575 1580

gat gca tgc tct ctg cca tct gga tcg gtt gga act cag tct gac cta 5103
Asp Ala Cys Ser Leu Pro Ser Gly Ser Val Gly Thr Gln Ser Asp Leu
1585 1590 1595

gga gca aac att gag ggt caa aat gtc aca aca gtg gct caa ctt ccc 5151
Gly Ala Asn Ile Glu Gly Gln Asn Val Thr Thr Val Ala Gln Leu Pro
1600 1605 1610

aca gat gga tca gat gca gtt gta acc ggt gga tct cct gta tca gat 5199
Thr Asp Gly Ser Asp Ala Val Val Thr Gly Gly Ser Pro Val Ser Asp
1615 1620 1625 1630

cag tgt gcc cag gat gca tct cct atg cca tta tct tcg cct gga aat 5247
Gln Cys Ala Gln Asp Ala Ser Pro Met Pro Leu Ser Ser Pro Gly Asn
1635 1640 1645

cac cct gat aca gca gtt aat atc gag ggt tta gat aac aca tca gta 5295
His Pro Asp Thr Ala Val Asn Ile Glu Gly Leu Asp Asn Thr Ser Val
1650 1655 1660

gct gag cct cat ata agt gga tca gat gca tgt gaa atg gaa att tca 5343
Ala Glu Pro His Ile Ser Gly Ser Asp Ala Cys Glu Met Glu Ile Ser
1665 1670 1675

gaa cct ggt ccc caa gta gag cgg tca acc ttt gca aat ctt ttc cat 5391
Glu Pro Gly Pro Gln Val Glu Arg Ser Thr Phe Ala Asn Leu Phe His
1680 1685 1690

gaa ggt ggc gtg gag cat tca gca ggt gta aca gct ctt gtt cca tca 5439
Glu Gly Gly Val Glu His Ser Ala Gly Val Thr Ala Leu Val Pro Ser
1695 1700 1705 1710

ctt ctt aac aat ggt acg gaa cag att gcc gtt caa cct gtt cct caa 5487
Leu Leu Asn Asn Gly Thr Glu Gln Ile Ala Val Gln Pro Val Pro Gln
1715 1720 1725

ata cct ttc cct gtg ttc aac gac ccg ttt ctg cat gaa ctg gag aag 5535
Ile Pro Phe Pro Val Phe Asn Asp Pro Phe Leu His Glu Leu Glu Lys

```
        1730                1735                1740

ttg cgg aga gaa tca gag aac tca aag aag act ttt gaa gaa aaa aaa    5583
Leu Arg Arg Glu Ser Glu Asn Ser Lys Lys Thr Phe Glu Glu Lys Lys
    1745                1750                1755

tca atc ttg aaa gct gaa ctc gag agg aag atg gct gaa gta caa gca    5631
Ser Ile Leu Lys Ala Glu Leu Glu Arg Lys Met Ala Glu Val Gln Ala
  1760                1765                1770

gag ttt cga aga aaa ttt cat gag gta gaa gcc gag cat aac acc aga    5679
Glu Phe Arg Arg Lys Phe His Glu Val Glu Ala Glu His Asn Thr Arg
1775                1780                1785                1790

acg aca aag ata gag aag gat aag aat ctt gtt ata atg aac aaa ctg    5727
Thr Thr Lys Ile Glu Lys Asp Lys Asn Leu Val Ile Met Asn Lys Leu
            1795                1800                1805

ttg gcg aat gcg ttc ttg tcc aaa tgt act gac aag aag gta tct ccc    5775
Leu Ala Asn Ala Phe Leu Ser Lys Cys Thr Asp Lys Lys Val Ser Pro
        1810                1815                1820

tca gga gct cca agg ggt aaa att cag cag cta gca cag aga gca gca    5823
Ser Gly Ala Pro Arg Gly Lys Ile Gln Gln Leu Ala Gln Arg Ala Ala
    1825                1830                1835

caa gtg agt gca ctg aga aat tac att gct cct cag cag ctt cag gca    5871
Gln Val Ser Ala Leu Arg Asn Tyr Ile Ala Pro Gln Gln Leu Gln Ala
  1840                1845                1850

tct tct ttt cct gct cct gct ctg gtt tcg gct cct ctg caa ctt cag    5919
Ser Ser Phe Pro Ala Pro Ala Leu Val Ser Ala Pro Leu Gln Leu Gln
1855                1860                1865                1870

caa tca tca ttt cct gct cct ggt ccg gct cct ctg cag cct cag gca    5967
Gln Ser Ser Phe Pro Ala Pro Gly Pro Ala Pro Leu Gln Pro Gln Ala
            1875                1880                1885

tct tcg ttt cct tct tca gtc tct cgt cca tca gcc ctt ctt ctg aat    6015
Ser Ser Phe Pro Ser Ser Val Ser Arg Pro Ser Ala Leu Leu Leu Asn
        1890                1895                1900

ttt gcg gtc tgt cca atg cct cag ccc aga cag cct ctc ata tcc aac    6063
Phe Ala Val Cys Pro Met Pro Gln Pro Arg Gln Pro Leu Ile Ser Asn
    1905                1910                1915

ata gct cca act cca tca gtt act cct gca aca aat cca ggt ctg cgt    6111
Ile Ala Pro Thr Pro Ser Val Thr Pro Ala Thr Asn Pro Gly Leu Arg
  1920                1925                1930

tct cct gca cca cac cta aac tca tat aga cca tcc tct tca act ccc    6159
Ser Pro Ala Pro His Leu Asn Ser Tyr Arg Pro Ser Ser Ser Thr Pro
1935                1940                1945                1950

gtc gcc aca gct act cca acc tcg tca gtg cct cct caa gct ttg aca    6207
```

Val Ala Thr Ala Thr Pro Thr Ser Ser Val Pro Pro Gln Ala Leu Thr
1955 1960 1965

tat tca gct gtg tca att cag cag cag caa gaa caa caa ccg caa cag 6255
Tyr Ser Ala Val Ser Ile Gln Gln Gln Gln Glu Gln Gln Pro Gln Gln
1970 1975 1980

agc ttg agc agt gga ttg cag agc aac aat gaa gtg gtt tgt ctt tct 6303
Ser Leu Ser Ser Gly Leu Gln Ser Asn Asn Glu Val Val Cys Leu Ser
1985 1990 1995

gac gac gag tgacctaaga ggagagatgg ttagggtctt agttattgat 6352
Asp Asp Glu
2000

ttttagagag ttaataatag tatatatata tatgtataag taggttacct aatctctgtc 6412

gttaatctaa tttagtgagt caggaaccga ctcgttggct aaggtctctc cttttgaaac 6472

gcaacgttct actttcatgt atataaatac agtctgatca cacaacacaa attgatgatt 6532

gaaaatacta ctgatttaac ttaaaaaaaa aaaaaaaaa 6571

<210> 3
<211> 2001
<212> PRT
<213> Arabidopsis thaliana

<400> 3
Met Lys Lys Asp Glu Lys Ile Gly Leu Thr Gly Arg Thr Ile Tyr Thr
1 5 10 15

Arg Ser Leu Ala Ala Ser Ile Pro Ala Ser Val Glu Gln Glu Thr Pro
20 25 30

Gly Leu Arg Arg Ser Ser Arg Gly Thr Pro Ser Thr Lys Val Ile Thr
35 40 45

Pro Ala Ser Ala Thr Arg Lys Ser Glu Arg Leu Ala Pro Ser Pro Ala
50 55 60

Ser Val Ser Lys Lys Ser Gly Gly Ile Val Lys Asn Ser Thr Pro Ser
65 70 75 80

Ser Leu Arg Arg Ser Asn Arg Gly Lys Thr Glu Val Ser Leu Gln Ser
85 90 95

Ser Lys Gly Ser Asp Asn Ser Ile Arg Lys Gly Asp Thr Ser Pro Asp
100 105 110

Ile Glu Gln Arg Lys Asp Ser Val Glu Glu Ser Thr Asp Lys Ile Lys
115 120 125

Pro Ile Met Ser Ala Arg Ser Tyr Arg Ala Leu Phe Arg Gly Lys Leu

130 135 140

Lys Glu Ser Glu Ala Leu Val Asp Ala Ser Pro Asn Glu Glu Glu Leu
145 150 155 160

Val Val Val Gly Cys Ser Arg Arg Ile Pro Ala Gly Asn Asp Asp Val
165 170 175

Gln Gly Lys Thr Asp Cys Pro Pro Pro Ala Asp Ala Gly Ser Lys Arg
180 185 190

Leu Pro Val Asp Glu Thr Ser Leu Asp Lys Gly Thr Asp Phe Pro Leu
195 200 205

Lys Ser Val Thr Glu Thr Glu Lys Ile Val Leu Asp Ala Ser Pro Ile
210 215 220

Val Glu Thr Gly Asp Asp Ser Val Ile Gly Ser Pro Ser Glu Asn Leu
225 230 235 240

Glu Thr Gln Lys Leu Gln Asp Gly Lys Thr Asp Cys Ser Pro Pro Ala
245 250 255

Asn Ala Glu Ser Lys Thr Leu Pro Val Gly Glu Thr Ser Leu Glu Lys
260 265 270

Glu Tyr Pro Gln Lys Phe Gln Asp Asp Asn Thr Asp Cys Leu Pro Pro
275 280 285

Ala Asn Ala Glu Ser Lys Arg Leu Pro Val Gly Glu Thr Ser Leu Glu
290 295 300

Lys Asp Thr Asp Phe Pro Leu Lys Ser Thr Thr Glu Thr Gly Lys Met
305 310 315 320

Val Leu Tyr Ala Ser Pro Ile Val Glu Thr Arg Asp Asp Ser Val Ile
325 330 335

Cys Ser Pro Ser Thr Asn Leu Glu Thr Gln Lys Leu Leu Val Ser Lys
340 345 350

Thr Gly Leu Glu Thr Asp Ile Val Leu Pro Leu Lys Arg Lys Arg Asp
355 360 365

Thr Ala Glu Ile Glu Leu Asp Ala Cys Ala Thr Val Ala Asn Gly Asp
370 375 380

Asp His Val Met Ser Ser Asp Gly Val Ile Pro Ser Pro Ser Gly Cys
385 390 395 400

Lys Asn Asp Asn Arg Pro Glu Met Cys Asn Thr Cys Lys Lys Arg Gln
405 410 415

Lys Val Asn Gly Asp Cys Gln Asn Arg Ser Val Cys Ser Cys Ile Val
420 425 430

Gln Pro Val Glu Glu Ser Asp Asn Val Thr Gln Asp Met Lys Glu Thr
435 440 445

Gly Pro Val Thr Ser Arg Glu Tyr Glu Glu Asn Gly Gln Ile Gln His
450 455 460

Gly Lys Ser Ser Asp Pro Lys Phe Tyr Ser Ser Val Tyr Pro Glu Tyr
465 470 475 480

Trp Val Pro Val Gln Leu Ser Asp Val Gln Leu Glu Gln Tyr Cys Gln
485 490 495

Thr Leu Phe Ser Lys Ser Leu Ser Leu Ser Ser Leu Ser Lys Ile Asp
500 505 510

Leu Gly Ala Leu Glu Glu Thr Leu Asn Ser Val Arg Lys Thr Cys Asp
515 520 525

His Pro Tyr Val Met Asp Ala Ser Leu Lys Gln Leu Leu Thr Lys Asn
530 535 540

Leu Glu Leu His Glu Ile Leu Asp Val Glu Ile Lys Ala Ser Gly Lys
545 550 555 560

Leu His Leu Leu Asp Lys Met Leu Thr His Ile Lys Lys Asn Gly Leu
565 570 575

Lys Ala Val Val Phe Tyr Gln Ala Thr Gln Thr Pro Glu Gly Leu Leu
580 585 590

Leu Gly Asn Ile Leu Glu Asp Phe Val Gly Gln Arg Phe Gly Pro Lys
595 600 605

Ser Tyr Glu His Gly Ile Tyr Ser Ser Lys Lys Asn Ser Ala Ile Asn
610 615 620

Asn Phe Asn Lys Glu Ser Gln Cys Cys Val Leu Leu Leu Glu Thr Arg
625 630 635 640

Ala Cys Ser Gln Thr Ile Lys Leu Leu Arg Ala Asp Ala Phe Ile Leu
645 650 655

Phe Gly Ser Ser Leu Asn Pro Ser His Asp Val Lys His Val Glu Lys
660 665 670

Ile Lys Ile Glu Ser Cys Ser Glu Arg Thr Lys Ile Phe Arg Leu Tyr
675 680 685

Ser Val Cys Thr Val Glu Glu Lys Ala Leu Ile Leu Ala Arg Gln Asn
690 695 700

Met Arg Gln Asn Lys Ala Val Glu Asn Leu Asn Arg Ser Leu Thr His
705 710 715 720

Ala Leu Leu Met Trp Gly Ala Ser Tyr Leu Phe Asp Lys Leu Asp His
725 730 735

Phe His Ser Ser Glu Thr Pro Asp Ser Gly Val Ser Phe Glu Gln Ser
740 745 750

Ile Met Asp Gly Val Ile His Glu Phe Ser Ser Ile Leu Ser Ser Lys
755 760 765

Gly Gly Glu Glu Asn Glu Val Lys Leu Cys Leu Leu Leu Glu Ala Lys
770 775 780

His Ala Gln Gly Thr Tyr Ser Ser Asp Ser Thr Leu Phe Gly Glu Asp
785 790 795 800

His Ile Lys Leu Ser Asp Glu Glu Ser Pro Asn Ile Phe Trp Ser Lys
805 810 815

Leu Leu Gly Gly Lys Asn Pro Met Trp Lys Tyr Pro Ser Asp Thr Pro
820 825 830

Gln Arg Asn Arg Lys Arg Val Gln Tyr Phe Glu Gly Ser Glu Ala Ser
835 840 845

Pro Lys Thr Gly Asp Gly Gly Asn Ala Lys Lys Arg Lys Lys Ala Ser
850 855 860

Asp Asp Val Thr Asp Pro Arg Val Thr Asp Pro Pro Val Asp Asp Asp
865 870 875 880

Glu Arg Lys Ala Ser Gly Lys Asp His Met Gly Ala Leu Glu Ser Pro
885 890 895

Lys Val Ile Thr Leu Gln Ser Ser Cys Lys Ser Ser Gly Thr Asp Gly
900 905 910

Thr Leu Asp Gly Asn Asp Ala Phe Gly Leu Tyr Ser Met Gly Ser His
915 920 925

Ile Ser Gly Ile Pro Glu Asp Met Leu Ala Ser Gln Asp Trp Gly Lys
930 935 940

Ile Pro Asp Glu Ser Gln Arg Arg Leu His Thr Val Leu Lys Pro Lys
945 950 955 960

Met Ala Lys Leu Cys Gln Val Leu His Leu Ser Asp Ala Cys Thr Ser
965 970 975

Met Val Gly Asn Phe Leu Glu Tyr Val Ile Glu Asn His Arg Ile Tyr
980 985 990

Glu Glu Pro Ala Thr Thr Phe Gln Ala Phe Gln Ile Ala Leu Ser Trp
995 1000 1005

Ile Ala Ala Leu Leu Val Lys Gln Ile Leu Ser His Lys Glu Ser Leu

1010 1015 1020

Val Arg Ala Asn Ser Glu Leu Ala Phe Lys Cys Ser Arg Val Glu Val
025 1030 1035 1040

Asp Tyr Ile Tyr Ser Ile Leu Ser Cys Met Lys Ser Leu Phe Leu Glu
1045 1050 1055

His Thr Gln Gly Leu Gln Phe Asp Cys Phe Gly Thr Asn Ser Lys Gln
1060 1065 1070

Ser Val Val Ser Thr Lys Leu Val Asn Glu Ser Leu Ser Gly Ala Thr
1075 1080 1085

Val Arg Asp Glu Lys Ile Asn Thr Lys Ser Met Arg Asn Ser Ser Glu
1090 1095 1100

Asp Glu Glu Cys Met Thr Glu Lys Arg Cys Ser His Tyr Ser Thr Ala
105 1110 1115 1120

Thr Arg Asp Ile Glu Lys Thr Ile Ser Gly Ile Lys Lys Lys Tyr Lys
1125 1130 1135

Lys Gln Val Gln Lys Leu Val Gln Glu His Glu Glu Lys Lys Met Glu
1140 1145 1150

Leu Leu Asn Met Tyr Ala Asp Lys Lys Gln Lys Leu Glu Thr Ser Lys
1155 1160 1165

Ser Val Glu Ala Ala Val Ile Arg Ile Thr Cys Ser Arg Thr Ser Thr
1170 1175 1180

Gln Val Gly Asp Leu Lys Leu Leu Asp His Asn Tyr Glu Arg Lys Phe
185 1190 1195 1200

Asp Glu Ile Lys Ser Glu Lys Asn Glu Cys Leu Lys Ser Leu Glu Gln
1205 1210 1215

Met His Glu Val Ala Lys Lys Lys Leu Ala Glu Asp Glu Ala Cys Trp
1220 1225 1230

Ile Asn Arg Ile Lys Ser Trp Ala Ala Lys Leu Lys Val Cys Val Pro
1235 1240 1245

Ile Gln Ser Gly Asn Asn Lys His Phe Ser Gly Ser Ser Asn Ile Ser
1250 1255 1260

Gln Asn Ala Pro Asp Val Gln Ile Cys Asn Asn Ala Asn Val Glu Ala
265 1270 1275 1280

Thr Tyr Ala Asp Thr Asn Cys Met Ala Ser Lys Val Asn Gln Val Pro
1285 1290 1295

Glu Ala Glu Asn Thr Leu Gly Thr Met Ser Gly Gly Ser Thr Gln Gln
1300 1305 1310

Val His Glu Met Val Asp Val Arg Asn Asp Glu Thr Met Asp Val Ser
1315 1320 1325

Ala Leu Ser Arg Glu Gln Leu Thr Lys Ser Gln Ser Asn Glu His Ala
1330 1335 1340

Ser Ile Thr Val Pro Glu Ile Leu Ile Pro Ala Asp Cys Gln Glu Glu
345 1350 1355 1360

Phe Ala Ala Leu Asn Val His Leu Ser Glu Asp Gln Asn Cys Asp Arg
1365 1370 1375

Ile Thr Ser Ala Ala Ser Asp Glu Asp Val Ser Ser Arg Val Pro Glu
1380 1385 1390

Val Ser Gln Ser Leu Glu Asn Leu Ser Ala Ser Pro Glu Phe Ser Leu
1395 1400 1405

Asn Arg Glu Glu Ala Leu Val Thr Thr Glu Asn Arg Arg Thr Ser His
1410 1415 1420

Val Gly Phe Asp Thr Asp Asn Ile Leu Asp Gln Gln Asn Arg Glu Asp
425 1430 1435 1440

Cys Ser Leu Asp Gln Glu Ile Pro Asp Glu Leu Ala Met Pro Val Gln
1445 1450 1455

His Leu Ala Ser Val Val Glu Thr Arg Gly Ala Ala Glu Ser Asp Gln
1460 1465 1470

Tyr Gly Gln Asp Ile Cys Pro Met Pro Ser Ser Leu Ala Gly Lys Gln
1475 1480 1485

Pro Asp Pro Ala Ala Asn Thr Glu Ser Glu Asn Leu Glu Glu Ala Ile
1490 1495 1500

Glu Pro Gln Ser Ala Gly Ser Glu Thr Val Glu Thr Thr Asp Phe Ala
505 1510 1515 1520

Ala Ser His Gln Gly Asp Gln Val Thr Cys Pro Leu Leu Ser Ser Pro
1525 1530 1535

Thr Gly Asn Gln Pro Ala Pro Glu Ala Asn Ile Glu Gly Gln Asn Ile
1540 1545 1550

Asn Thr Ser Ala Glu Pro His Val Ala Gly Pro Asp Ala Val Glu Ser
1555 1560 1565

Gly Asp Tyr Ala Val Ile Asp Gln Glu Thr Met Gly Ala Gln Asp Ala
1570 1575 1580

Cys Ser Leu Pro Ser Gly Ser Val Gly Thr Gln Ser Asp Leu Gly Ala
585 1590 1595 1600

Asn Ile Glu Gly Gln Asn Val Thr Thr Val Ala Gln Leu Pro Thr Asp
1605 1610 1615

Gly Ser Asp Ala Val Val Thr Gly Gly Ser Pro Val Ser Asp Gln Cys
1620 1625 1630

Ala Gln Asp Ala Ser Pro Met Pro Leu Ser Ser Pro Gly Asn His Pro
1635 1640 1645

Asp Thr Ala Val Asn Ile Glu Gly Leu Asp Asn Thr Ser Val Ala Glu
1650 1655 1660

Pro His Ile Ser Gly Ser Asp Ala Cys Glu Met Glu Ile Ser Glu Pro
665 1670 1675 1680

Gly Pro Gln Val Glu Arg Ser Thr Phe Ala Asn Leu Phe His Glu Gly
1685 1690 1695

Gly Val Glu His Ser Ala Gly Val Thr Ala Leu Val Pro Ser Leu Leu
1700 1705 1710

Asn Asn Gly Thr Glu Gln Ile Ala Val Gln Pro Val Pro Gln Ile Pro
1715 1720 1725

Phe Pro Val Phe Asn Asp Pro Phe Leu His Glu Leu Glu Lys Leu Arg
1730 1735 1740

Arg Glu Ser Glu Asn Ser Lys Lys Thr Phe Glu Glu Lys Lys Ser Ile
745 1750 1755 1760

Leu Lys Ala Glu Leu Glu Arg Lys Met Ala Glu Val Gln Ala Glu Phe
1765 1770 1775

Arg Arg Lys Phe His Glu Val Glu Ala Glu His Asn Thr Arg Thr Thr
1780 1785 1790

Lys Ile Glu Lys Asp Lys Asn Leu Val Ile Met Asn Lys Leu Leu Ala
1795 1800 1805

Asn Ala Phe Leu Ser Lys Cys Thr Asp Lys Lys Val Ser Pro Ser Gly
1810 1815 1820

Ala Pro Arg Gly Lys Ile Gln Gln Leu Ala Gln Arg Ala Ala Gln Val
825 1830 1835 1840

Ser Ala Leu Arg Asn Tyr Ile Ala Pro Gln Gln Leu Gln Ala Ser Ser
1845 1850 1855

Phe Pro Ala Pro Ala Leu Val Ser Ala Pro Leu Gln Leu Gln Gln Ser
1860 1865 1870

Ser Phe Pro Ala Pro Gly Pro Ala Pro Leu Gln Pro Gln Ala Ser Ser
1875 1880 1885

Phe Pro Ser Ser Val Ser Arg Pro Ser Ala Leu Leu Leu Asn Phe Ala

1890 1895 1900

Val Cys Pro Met Pro Gln Pro Arg Gln Pro Leu Ile Ser Asn Ile Ala
905 1910 1915 1920

Pro Thr Pro Ser Val Thr Pro Ala Thr Asn Pro Gly Leu Arg Ser Pro
1925 1930 1935

Ala Pro His Leu Asn Ser Tyr Arg Pro Ser Ser Ser Thr Pro Val Ala
1940 1945 1950

Thr Ala Thr Pro Thr Ser Ser Val Pro Pro Gln Ala Leu Thr Tyr Ser
1955 1960 1965

Ala Val Ser Ile Gln Gln Gln Gln Glu Gln Gln Pro Gln Gln Ser Leu
1970 1975 1980

Ser Ser Gly Leu Gln Ser Asn Asn Glu Val Val Cys Leu Ser Asp Asp
985 1990 1995 2000

Glu

<210> 4
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Oligonucleotide

<400> 4
catctacggc aatgtaccag c 21

<210> 5
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Oligonucleotide

<400> 5
gatgggaatt ggctgagtgg c 21

<210> 6
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 6
cagttccaaa cgtaaaacgg c 21

<210> 7
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 7
ntcgastwts gwgtt 15

<210> 8
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 8
ngtcgaswga nawgaa 16

<210> 9
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 9
wgtgnagwan canaga 16

<210> 10
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic

Oligonucleotide

<400> 10
wggwancwga wangca 16

<210> 11
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 11
wcgwwgawca ngncga 16

<210> 12
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 12
wgcnagtnag wanaag 16

<210> 13
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 13
awgcangncw ganata 16

<210> 14
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 14

ctgtacatac tgagtacaat cgga 24

<210> 15
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 15
gcttcaattc ctgcctcagt tgaac 25

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 16
ctctacgtgc ttaacatcat gcga 24

<210> 17
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 17
ccagcttctg ctactagaaa gtcag 25

<210> 18
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 18
ctggagttgc atgaaatcct ggatg 25

<210> 19
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic Oligonucleotide

<400> 19
gctctttgta agctgttcac gagac 25

<210> 20
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic Oligonucleotide

<400> 20
tcgcatgatg ttaagcacgt agag 24

<210> 21
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic Oligonucleotide

<400> 21
gagtactggt ccgtgaacag gtaat 25

<210> 22
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic Oligonucleotide

<400> 22
atgcttgcac aagcatggtc ggaaa 25

<210> 23
<211> 25
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 23
tgcaacatcg tgcatttgct ccaga 25

<210> 24
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 24
cacaagcatg agtttttcct tccgg 25

<210> 25
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:Synthetic
Oligonucleotide

<400> 25
ctgactttct agtagcagaa gctgg 25

<210> 26
<211> 519
<212> DNA
<213> Brassica oleracea

<220>
<223> seq1-23

<400> 26
gaattcctgn nttacggcat ccttaataga ctgttcaaat ggaactcctg aacctggggt 60
tccactccca tggaagtgtt ccagcttatc aaataaatat gatgcccccc acatgagcaa 120
tgcatgtgtg agaggacggt ttaggttctc tagaggctta ttttgcctag caagaatcag 180
ggttttttct tcaactgtaa acactgagta caaccggaaa atcttagttc tttcagaaca 240
cgactcaacc tttatcttct ctaagagctt aacgtcatgc gatggattca ggctgcttcc 300
aaaaagtata aaagactcag cgcgtaagag tttaatgctt tgactacagg cacgtatttc 360
cagcagcaga ataaaacact cactctcctt gttgaaattg tttatagcgt tcttcttcga 420
gaggcagacc ccatgctcat aggaattttg accaaatctt tgcatcagaa aatcttcgag 480
aatattacca agcagaagcc cctcagggct atgtattgc 519

<210> 27
<211> 419
<212> DNA
<213> Brassica oleracea

<220>
<223> seq1-27

<400> 27
gaattcagga tcaaaagggt tgccggttgg agaaactggt ttagagaaag gctctgattt 60
tcctgtggaa gtaactaagg atatagagaa gacagtggtt gattcatccc ccatggttga 120
aactgaggat ggcagtgtta taggttcacc atccgagaat ccagaaccac aaaagcttcg 180
tgacagtgaa actagcttgg aaaccgatat agacttggct ctgaaaagaa aaagagacac 240
tgcagaaatt gtgatggatg catgtacaaa tgcagatgac cgcattatga gtactgatgg 300
ggttattcct tttccacccg tgtgcacaaa tattaatcaa cccgaaaggt gtggcacatg 360
tcaaaaacgg caaaagtaag aatttccgac tgttgtctgt cgttttgaaa ccatttgcc 419

<210> 28
<211> 467
<212> DNA
<213> Brassica oleracea

<220>
<223> seq1-43

<400> 28
gaattctcgt ccatactttc ttccgatgtt ggagaagaaa atgaaggcaa gctgtgtcta 60
cttttggaag ccaagcatgc tcagggaagt tacagcactg atgctactct atttggtgaa 120
gaacatgtca agttatcaga tgaaagtcca aatatgtttt ggtcaaagct gttgagtgga 180
aagaacccta tgtggaaata ctgttcggat actcctcaaa ggagtcgaaa aagagtacgg 240
catcttcagg gctatgagga gactaccaaa gttggcaatg gcggaaactt aaagaagaaa 300
aagaaggctt cagatgatgt cacagtagat aacgctgaga gaaaagcctc tggaaaggat 360
cacatgggta aaacagttca cttcctgctc ctttacctct agtgttcatt gaatgttcca 420
tttactttgc ttactatctt tccttcaggg catttggagt caccaaa 467

<210> 29
<211> 490
<212> DNA
<213> Brassica oleracea

<220>
<223> seq1-47

<400> 29
gaattcagct tttaaaactg atctctgctc acagataatt taagagtcag tgaaaattga 60
gataaaacga accaaaactg gaggtaacag atactctgag aacaactaac cttttcttca 120
taagtcttct ttgtgttctc tgattctctc cgcagcttct ccagttcatg ctgaaatggg 180
tcactgaaca cagggaaagg tacttgagga acaggtggag tggcattctg tcccgtagca 240
ttgttaagct gtgaagaaac aggagctgtt acacctgctg gaggctccac aacaccttca 300
tcgacaacgt ctgcgtaaaa ggtattacca gattgtcagt ttctctggca aacacatacg 360
ttatacttaa atgcaaaaga gcagttactg acttgcaaag gttggttgtt ctacttgagc 420
atcaggttct gctacttcca tttcacatgc ttctgatcca gttgtgcgag gcgcagccat 480

tgttgtgttg 490

<210> 30
<211> 515
<212> DNA
<213> Brassica oleracea

<220>
<223> 2-33

<400> 30
tctagagaag aggtggatta tgtatattct tttctgtact gcatgaagag tctattcgtg 60
gggcgcacac aaggtttcca agaaaagggt gaagaatgca tggctgagaa aagaggtagc 120
cattatagct cagtaaccaa ggatgttgaa aagactatta gcgacatcaa aaagaaatgc 180
agtaagagcc tgcataagct tgtacaaacc ctcgaggaag aaaagatgga cctgatgaat 240
aggaatgctg tcaagaagca ggaacttcag aattgtaaaa aggtggaagc atcatttatt 300
cgtgtcacct attcaggtat aaatactcag agcttacatg atgctctcca acggctggaa 360
tgtacttttg aaagaaagtt tgatgatctc aaaggagagt tggatgaatg ccttgaaagt 420
ttagagcaaa taaacgaggc tggaaagaag aagttggctg aagatgaagc ctgttggatt 480
agtcggatag agaaatgggc acgagctgaa ttaag 515

<210> 31
<211> 574
<212> DNA
<213> Brassica oleracea

<220>
<223> seq2-37

<400> 31
tctagaccaa actattaaac gctaaacata agaagattag atcactcgtc atcagagaga 60
cagaccacat cattgctcct ctgcaatcca ctccccaagt tctgtggttg ttcttgctgc 120
tgaataaacg catttgaata tggtaaaggg ttggagatga gaggttgtct tggttgaggc 180
attgtgcagt acggagccga agcagtatga ttcctcagtg cgcttacttg tgttgctctc 240
tgtgctagct gctggattct aactggagaa agaaaaaaag aaaaaaaagg tgttattatg 300
acttcataac cttatatctt taaaaaacaa ttatgcttct attattcgaa cacttgccca 360
ttggagttgc tgctgaggaa tgagaggaga ttctgctcgt acatttagac aagaacgcac 420
tcgacaacag cttgttcttt ataacaagat tcttcctcgt ctgtaacttc gtctttctgg 480
ctgcatgtac agcttgtacc tcatgaaact ttctctgata ctcttcttgt aattcagcta 540
tcttcttctc gaatttagct ttcaagactg cttt 574

<210> 32
<211> 466
<212> DNA
<213> Brassica oleracea

<220>
<223> seq2-53

<400> 32
tctagattgt aattttaaat ttacaacaaa ttttgaaagg gtcagcgatg agtttgcaaa 60
tctccgtgtt tcctccagca ttgctcagcc agttcaagaa cctgatcact tggcacaggt 120

tggtttcttc ttgctttact ttggacacct gtttaatatt ggcctgtcaa atttacttat 180
ccttttactt ctaaactgca aattctggtc tgcattgcat tgtgatatga aggtatctgg 240
acccgcttca agcagagact atggggagga caggcagaat atgcaacaag ataaatcaca 300
tgaccgaaag ttgtcatcga tgtatccaga gtattgggtt ccagtgcagc tatcagatgt 360
acagatagag caatactgtc ggactctctt ctccaaatct tcatctcttt cttcgctgtc 420
gaggactgat cctgttcgag ctcttgaaca aactctcagt tctgta 466

<210> 33
<211> 417
<212> DNA
<213> Brassica oleracea

<220>
<223> seq2-57

<400> 33
tctagagcaa ttgaaaccta attccgattt tgcgcgggcc agagattctt cacggttgaa 60
cttttgctta acgaaagaga ctgcaatcca aatctggaag tgcattatta agaacgtatt 120
cagcaatatt cataaattat gcaacaatca aaggccttac gttgtggcct acaaagcatg 180
gattttgtta gatattagta gctagtctaa ttcaagcaat taatggaagt ttctatccta 240
tgactggaaa gttaaacatt cccacaaaag cagtgatgcc acagatgatg aagaagaaaa 300
atgcatatac tatggaagtg aatgctatca taccacagct atctggaagg cctgcaatgt 360
tgtagctggc tctttgcaga cacggtggtt gtcaataata tattcaagaa ctttttc 417